Advances in psychotherapy research and precision mental health:

Answering the “What works for whom” question for patients with depression

Cumulative Doctoral Thesis presented to the faculty of Human Sciences

of the University of Bern

by M. Sc. Nadine Friedl

Accepted in the spring semester 2020

on the recommendation of the doctoral committee composed of

Prof. Dr. phil. Thomas Berger (main supervisor)

Prof. em. Dr. phil. Franz Caspar

Bern, 2020

Nothing in life is to be feared, it is only to be understood.

Now is the time to understand more, so that we may fear less. — Marie Curie

Acknowledgements

First of all, I would like to thank my supervisors Prof. Dr. phil. Thomas Berger and Prof. em. Dr. phil. Franz Caspar for their constant support and trust. I want to thank them for their guidance through each stage of the process. It was an inspiration to work with them as each one of them defined my path of research in one way or the other. I would like to acknowledge Thomas Berger for inspiring my interest in E-Mental Health and internet-based therapy, and Franz Caspar for inspiring my interest in the Generic Model of Psychotherapy and process-outcome research.

I would also like to thank Prof. Dr. phil. Martin Grosse Holtforth for the cooperation on one of the projects. He gave me a great opportunity that shaped my expertise and I feel grateful to have learned so much from him.

Furthermore, I would like to thank Dr. phil. Tobias Krieger who patiently supported me on every single manuscript and who's door was always open. His support meant a lot to me, not only as a researcher but also as a psychotherapist and mentor in the University outpatient center where I am training to become a psychotherapist.

A special thank you goes to my colleagues, especially Nina Rüegg, Christian Moser, Laura Bielinski and everyone in "the corner office". I am forever grateful for their emotional and moral support and all the good laughs. It would not have been possible without them.

Last but not least, I would like to thank my family. My parents, because without them, I would not be where I am today. My grandparents, for always being there for me. My grandmother *Oma Steffi,* who left me way too early and who would have loved to see this. But I know I made her proud.

Summary

The present doctoral thesis focuses on two articles which are embedded in the field of precision mental health and treatment selection. Study 1 examined if model determined treatment allocation to cognitive behavioral therapy (CBT) or CBT with integrated exposure and emotion-focused elements (CBT-EE) results in better treatment outcomes while using important predictors found for each intervention. Study 2 investigated important predictors in routine care and blended internet- and face-to-face CBT in secondary care, as well as treatment outcomes for treatment allocation using this predictive information. Both studies use a Bayesian approach called Bayesian Model Averaging (BMA) and the Personalized Advantage Index (PAI) for their statistical analyses. After an introduction to the Generic Model of Psychotherapy, the development of process and outcome research and the thematic field of treatment selection and precision medicine, the individual articles will be described and critically reflected in more detail. Possibilities and limits of predicting the optimal treatment for an individual based on algorithms are discussed based on the results of the two studies. Taken together, the two studies provide an important contribution to psychotherapy research as the feasibility of treatment selection using BMA and PAI is shown. Last but not least, implications for future research are discussed and an example of how treatment selection can be transferred into clinical practice is presented.

Content

1 General introduction

Almost half of all people around the globe meet the criteria of a clinically relevant mental disorder at least once in their lives. Approximately one in three suffer from an anxiety disorder once in their lives and almost one in six suffer from a depressive episode (Kessler et al., 2005). According to the World Health Organization, 300 million people globally suffer from major depressive disorder (MDD) and by 2030 it will be one of the leading causes of the global burden of disease (Mathers & Loncar, 2006; World Health Organization, 2018). It is one of the most common problems seen in clinical practice and it is associated with a great deal of suffering and societal costs (World Health Organization, 2008). One reason for this is the chronic-intermittent course that MDD and anxiety disorders can take resulting in a sustained impact on the quality of life (Penninx et al., 2011). Furthermore, it is assumed that up to 20-25% of MDD patients are at risk for chronic depression (Dinga et al., 2018).

In order to deal with the increasing prevalence of mental disorders worldwide, appropriate treatment options are needed. Well established treatment approaches for MDD are cognitive behavioral therapy (CBT), interpersonal psychotherapy (IPT) and antidepressant medication (ADM). Although these approaches are well studied and have shown to be efficacious (Cuijpers et al., 2013; Hollon, Cohen, Singla, & Andrews, 2019), a high percentage of people that suffer from MDD or anxiety disorders are either not in treatment or do not benefit from it as much as expected (Cohen & DeRubeis, 2018). As a result, there are low treatment response rates and high dropout rates in treatments for depression (Cuijpers & Christensen, 2017; Hollon et al., 2019; Lemmens et al., 2019). Different factors can be held accountable for this. First, accessibility and availability of adequate treatment options are limited (Cavanagh, 2014). Second, one needs to recognize that no single treatment is likely to be the best for everyone. Sophisticated and new statistical approaches like machine learning

and new technologies and advances in psychotherapy research have a great potential to improve overall response rates and accessibility of psychological treatments.

1.1 Process outcome research

Over the years, psychotherapy research as we know it now has come a long way and it continues to be characterized by new developments and changing emphases (Lambert, 2013). Historically, it underwent four developmental phases with each of them addressing a fundamental question. The legitimation phase and the second phase, the competition phase answer the questions “Is psychotherapy effective” and respectively “Which form of psychotherapy is better?”. The prescriptive psychotherapy phase addresses the question “Which form of psychotherapy is indicated for whom?” and the last phase is called the process-research phase which looks at the mechanism of change in psychotherapy (Grawe, 1997).

Although the studies that are being discussed in this thesis focus on the prescriptive phase, it is important to look at the bigger picture, namely the process-outcome research because it offers a potential framework on how to answer the fundamental question of the prescriptive phase. Process-outcome research is one field in psychotherapy research that measures process variables and tests whether they relate to therapy outcome (Hill & Corbett, 1993; Orlinsky & Howard, 1986; Orlinsky, Ronnestad, & Willutzki, 2004; Timulak, 2008). First, it is essential to define what is meant by process and outcome. *Process* can be defined as everything that happens in the therapy sessions, thus therapist and client behaviors but also the interaction between the two. This is only one possible definition of process because there is great variability in the meaning attached to it (Orlinsky et al., 2004). The *outcome* refers to the changes that occur as a result of the psychotherapeutic process (Hill & Corbett, 1993; Timulak, 2008). There is little consensus whether or not the outcome is the result at the end of

therapy or whether smaller outcomes in and between sessions should be defined as such as well (Greenberg & Pinsof, 1986).

Similar to the developmental process of psychotherapy research, process-outcome research developed through different stages as well (Orlinsky et al., 2004). The initial phase (1920-1950) had the goal to establish scientific research into psychotherapy (Bergin, 1971). In a second phase, treatment monitoring was established through the recording of therapy sessions. Since 1960, the use of diagnostic questionnaires and rating scales was established with the aim to install scientific accuracy (Orlinsky et al., 2004). The third historical phase of process-outcome research can be described as one of expansion, differentiation, and organization (Orlinsky & Russell, 1994). This phase is especially characterized by the first meta-analysis to demonstrate the effectiveness of psychotherapy (Smith, Glass, & Miller, 1980). According to this meta-analysis, 80% of patients are better off after treatment compared to an untreated sample (Hill & Corbett, 1993). The greatest accomplishment of this phase was the introduction of the effect size as the universal comparative parameter. This allowed for all researchers to integrate and compare their findings (Orlinsky et al., 2004). Moreover, there was a revival of interest in the working alliance concept which remains strong up to today (Horvath & Greenberg, 1994). The working alliance is still seen as one of the core contributions to a successful therapy outcome and is one of the most researched variables in psychotherapy research (Flückiger, Del Re, Wampold, & Horvath, 2018). Furthermore, the method “task analysis” as a way to track and analyze in-session events that can be used as markers for therapeutically significant events was introduced (Rice & Greenberg, 1984). The fourth and last phase of the development of process-outcome research started in around 1985 and was dominated by consolidation, standardization and elaboration. The results of several long-term studies were published and with regard to standardization, randomized controlled trials (RCTs) were established as the gold standard of psychotherapy

research (Bothwell, Greene, Podolsky, & Jones, 2016). Elaboration in this phase stands for the use of new statistical methods that allow for more sophisticated analyses (Kraemer & Thiemann, 1989). Moreover, this phase is characterized by critique, innovation and controversy. For example, the criticism of group comparison research designs led to a shift from treatment focused research to patient focused research (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Orlinsky et al., 2004). Furthermore, the use of qualitative methods was highly encouraged because the integration of quantitative and qualitative methods retains the richness of data (Hill & Corbett, 1993; Orlinsky et al., 2004). One can say that we are still in this last phase as the constructs named above still play an important role in process-outcome research today. To date, more than 2000 process outcome studies have been published including several comprehensive reviews (Llewelyn, Macdonald, & Doorn, 2016).

1.2 The Generic Model of Psychotherapy

With the growing number of process outcome studies, Orlinsky and Howard thought of a model that integrates the results of these studies called the “Generic Model of Psychotherapy” (Orlinsky & Howard, 1987). It is conceived as a transtheoretical frame for integrating various empirical findings and it offers a comprehensive framework in which clinical theories of psychotherapy can be compared and combined (Orlinsky, 2009). The Generic Model of Psychotherapy integrates 2354 findings of almost 500 psychotherapy studies that examine the relationship of a process variable with an outcome variable. The goal was to create a research theory or in other words a research-based metatheory of psychotherapy. It links clinical practice theories or treatment models independent of their theoretical background which ultimately leads to a great source of guidance for empirical investigations. Furthermore, it acts as a conceptual framework that guides psychotherapy researchers in the development and implementation of studies. Most importantly, the Generic Model of Psychotherapy formulates a general psychotherapy model that is independent of the

different approaches to psychotherapy (Orlinsky, 2009; Orlinsky & Howard, 1987; Orlinsky et al., 2004).

The Generic Model of Psychotherapy is shown in Figure 1. It consists of three levels of distinction. First, the *input* is the context in which psychotherapy takes place. It is important to acknowledge that psychotherapy is embedded in the social context of the patient. One cannot say the patient is in therapy but one has to say that the therapy takes place in the life of the patient and is thus a part of it. The variables included in the input have the potential to influence the process of psychotherapy. Second, the *process* is the key component of the model. It includes the interactions between the therapist and patient during therapy. The third level of distinction is the *output* which is the result of this process. The therapy has effects that go beyond the direct outcome for the patient. It has consequences for the environment of the patient, for the therapist and for the society as well (Orlinsky & Howard, 1987).

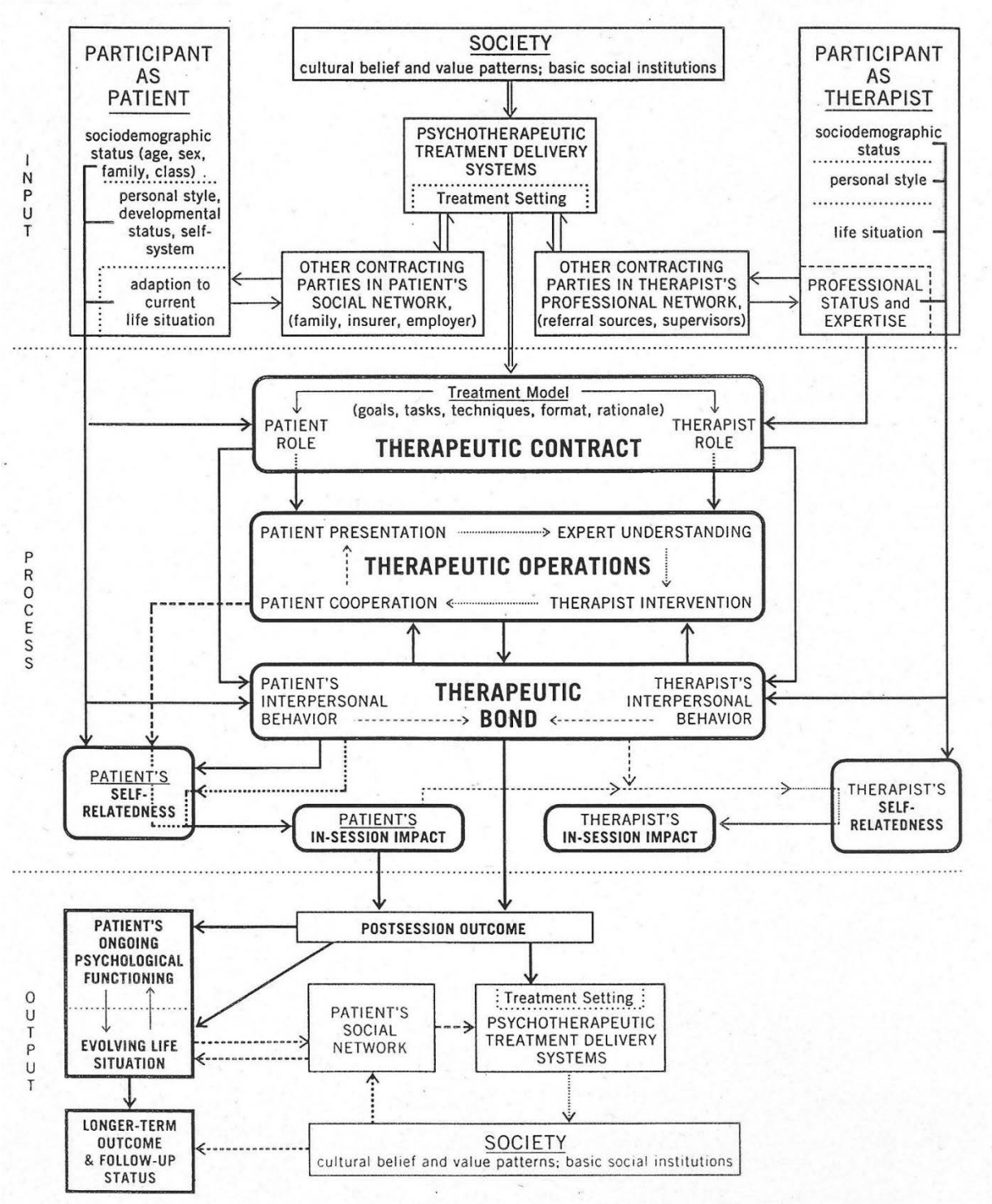

Figure 1. The Generic Model of Psychotherapy (Orlinsky & Howard, 1987).

The heart of the model, the *process* includes five active psychotherapeutic processes and the component *temporal patterns* which was added to the model in later years when research began to be interested in temporal patterns in therapy and their relation to output.

Each process category includes the role of the patient as well as the role of the therapist. The *therapeutic contract* represents the organizational aspect of therapy, e.g. therapy goals, the format and the rationale. It defines the norms of the participant's roles as therapist and patient (Orlinsky et al., 2004). The *therapeutic operations* represent the technical aspect of psychotherapy in the form of the cycle of reciprocal role-specific behavior. Patient presentation leads to the expert understanding and as a result to the therapist intervention. The intervention of the therapist in turn leads to the cooperation of the patient. The interpersonal aspect of therapy is defined as the *therapeutic bond* which describes the quality of the involvement of the therapist and the patient and the respective interpersonal behavior. It includes a socio-emotional component as well as the working alliance. The fourth process category is the therapist's and patient's *self-relatedness* which stands for their intrapersonal behavior. The clinical aspect of therapy is shown in the *in-session impacts*. In-session impacts are immediate positive or negative impacts resulting from the therapeutic operations. The sixth category *temporal patterns* represents the sequential aspect of process. This can include sequelae of events and moments during the sessions, periods or stages in the treatment stage or temporal aspects of the whole treatment course. The first five categories are concurrent facets of the psychotherapeutic process and cannot be seen as distinct stages, thus they do not occur in the order named above but they are in a constant interaction with each other. Each form of psychotherapy includes an individual configuration of the process categories (Orlinsky et al., 2004).

In the Generic Model of Psychotherapy, psychotherapy is viewed as a system of action that has individual and collective contexts that influence therapeutic processes in the form of input and, in turn, are influenced by therapeutic processes (output). Relationships of process variables and outcome variables that have shown to be robust findings included in the Generic Model are on the side of the patient (1) preparation of the patient role, (2) suitability of the

patient, (3) cognitive and behavioral processes, (4) cooperation, (5) positive affect, (6) engagement, (7) openness and (8) articulateness. Hence, the active participation of the patient in the therapy seems to play an important role for a successful therapy. For the therapist, variables that showed a consistent relationship with the output are (1) problem actualization and confrontation, (2) interpretation and (3) paradox interventions. Furthermore, the therapeutic bond, appreciation and the length of therapy had a robust relationship to the output (Kolden & Howard, 1992; Orlinsky et al., 2004).

To conclude, the results show that effective psychotherapy is more than a set of techniques but also more than a good therapeutic alliance. Moreover, the therapeutic outcome seems to be the result of the problems and resources of the patient in combination with the competencies and limits of the therapist. The Generic Model of Psychotherapy is a mean to help with the common issue of clinicians to respond to and integrate research findings into routine practice and for researchers to translate their findings into clinically useful recommendations for practice (Lambert, 2013).

1.3 Precision medicine and treatment selection

In Bern, Switzerland, there is a long tradition of differential psychotherapy research. As early as in the 1970s, Grawe was one of the first to consistently advocate differential psychotherapy, i.e. psychotherapy that takes into account important distinguishing characteristics of patients (Grawe, 1976). A couple of years later, in the beginning of the 90ies, the results of the Bernese comparative treatment study were published. Again, differential psychotherapy research played a major role (Grawe, Caspar, & Ambühl, 1990). Four different types of therapy have been compared and numerous research questions have been examined. As an example, it was found that for individual broad spectrum behavior therapy, patient’s dispositions before therapy correlated highly with treatment outcome. The

authors suggested that some patients may benefit from this therapy whereas others may not (Grawe et al., 1990).

On an individual level, differential responses to treatments are not a scarcity (Simon & Perlis, 2010) because patients with depression vary in their treatment response and illness course (Cuijpers, Van Straten, Bohlmeijer, Hollon, & Andersson, 2010). In addition, patient characteristics can moderate the efficacy of different treatments (Fournier et al., 2009). Therefore, it is important to recognize that no treatment is likely to be the best for everyone thus no particular treatment works universally, across all patients and most interventions work well on some patients (Beutler & Harwood, 2002; Cohen & DeRubeis, 2018). But by finding for each individual the treatment that works best for him or her, it may be possible to improve the overall treatment response rates (Hollon et al., 2019).

This is what guided Beutler and Harwood when they first introduced the term *Prescriptive Psychotherapy* and when they published their practical guide to systematic treatment selection (Beutler & Harwood, 2000). It is a detailed manual that provides a set of principles which can be used from multiple theoretical perspectives. Their goal was to be able to more effectively address the differences that exist between patients, since it is assumed that patients' reactions moderate treatment effects. *Prescriptive Psychotherapy* aims to match the setting, the patient and the therapist's dispositions with the patient's preferences. Its goal is to be able to make a clear treatment recommendation to the patient on which therapy is the best for him or her. According to Beutler and Harwood, the variables that interact most successfully with treatment are coping strategies, resistance, severity of the problem and the level of suffering (Beutler & Harwood, 2000). Furthermore, the term *Systematic Treatment Selection (STS)* is important to mention as it follows the ideas of *Prescriptive Psychotherapy* in the sense that it assumes that there is no treatment model that works well for all patients and that most treatment methods work well for some patients (Beutler & Clarkin, 1990;

Beutler, Clarkin, & Bongar, 2000; Beutler, Harwood, Bertoni, & Thomann, 2006). STS has the goal to identify individual dispositional factors and the interventions that they most effectively correspond with while it is constructed around principles of behavior change irrespective of different therapeutic perspectives (Beutler et al., 2000; Beutler et al., 2006).

Precision medicine has a long history of tailoring treatments to the specific needs and characteristics of a given patient (Katsnelson, 2013). It has shown to be especially successful in personalized treatment selection for cancer patients (Schwaederle et al., 2015). The use of algorithms to improve and support clinical decision-making in psychotherapy is growing as well (Cohen & DeRubeis, 2018; Hamburg & Collins, 2010). In fact, personalizing treatments for depression is one of the major challenges and promises for mental health research (Lopez-Gomez et al., 2019). With regard to precision medicine in psychotherapy research, the following question plays an important role because it guides today's efforts of personalizing treatments in psychotherapy: *"In all its complexity, the question towards which all outcome research should ultimately be directed is the following: What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?"* (Paul, 1967, p. 117). This question is already very old and the fact that we have had surprisingly little resilient insight into it so far illustrates how difficult it is to find answers.

So how do clinicians and researchers know which patients will benefit from which treatments? A trial and error approach is used nowadays to find each patient a treatment modality that is helpful for him or her (Cuijpers & Christensen, 2017). As an example, patients often go through multiple antidepressants before an effective therapy regimen is identified (Rush et al., 2006). So far, different attempts at personalizing treatments have been made. Pharmacogenetics tried to improve pharmacotherapies with genotyping as 50% of the response to antidepressants can be attributed to genetic factors (Cuijpers & Christensen, 2017;

Holsboer, 2008). Furthermore, efforts have been made to divide depression into different subtypes that might respond differentially to different treatments (Baumeister & Gordon, 2012). An approach that has become more and more popular is to compare two different treatments in an RCT and identify outcome moderators for each treatment while acknowledging individual patient differences (Kraemer, 2013). One of the earliest examples is the "matching factor" that combined the prescriptive value of different pre-treatment variables and the baseline symptom severity in a linear model that predicts symptom change (Barber & Muenz, 1996). Another example is the "nearest neighbors" technique where every patient's outcome in each treatment is predicted from the average observed post-treatment score of patients who are similar to the index patient (Lutz et al., 2006). Advances like that and the heterogeneity of statistical analyses have shown promising results that have the potential to enhance personalized medicine in psychotherapy (Cohen & DeRubeis, 2018).

1.4 The Personalized Advantage Index (PAI)

The PAI approach is a treatment selection approach that uses machine learning and multivariable regression modeling to offer a solution for the challenges and problems in precision mental health. It identifies patients with certain characteristics for whom one treatment works better than another (Cohen & DeRubeis, 2018; Cohen, Kim, Van, Dekker, & Driessen, 2019). It then provides a quantitative estimate of how much one particular treatment is better for this patient than for another. The PAI is a promising approach to increase the likelihood of improving a patient's mental health (DeRubeis, Cohen, et al., 2014).

This approach was first developed in 2011 by Robert DeRubeis and his team when they saw the results of an RCT that compared cognitive therapy (CT) with ADM for patients with depression. They found that on a group level, the treatments were equally effective but they also found variables that were associated with differential responses to treatment (DeRubeis et

al., 2005; Fournier et al., 2009). As an example, patients with comorbid personality disorder improved more with ADM than they did in CT whereas patients who were unemployed improved more in CT than in ADM. So how does a clinician use information from multiple conflicting predictors? According to early research, in cases like that, actuarial approaches to treatment selection seem to be superior to clinical judgement (Dawes, Faust, & Meehl, 1989). Others argue that clinical judgement makes a major contribution to therapy outcome and that it is essential in terms of constraints and resources of the patients (Caspar, 1997). The debate of which one is superior to the other cannot be answered definitively as strengths and difficulties can be found for both approaches (Meehl, 1954). The PAI may ultimately provide a solution for the challenge named above (Cohen & DeRubeis, 2018; DeRubeis, Cohen, et al., 2014).

The method can be used in any context where patients have been randomized to two or more treatment conditions that are equally effective on a group level. What is distinctive for the PAI approach is to identify pre-treatment variables that predict treatment outcome, either differential or for all treatments. In a next step, these variables are used to generate predictions for the treatment outcome of each patient for each treatment condition. As a result, each patient has a *factual prediction*, which is the treatment outcome for the treatment the patient actually received, and a *counterfactual prediction* for the treatment condition the patient did not receive. Thereupon, the prediction that indicates a better outcome is called the *optimal* treatment for that patient vs. the *suboptimal* treatment. The predicted difference of receiving the treatment with the greater predicted benefit versus the suboptimal treatment is an index of predicted advantage which is then called the PAI. Higher values of the PAI stand for a stronger predicted benefit (DeRubeis, Cohen, et al., 2014). When DeRubeis and his team first tested the PAI approach, they suggested to generate the predictions in the same model using a leave-one-out-cross validation. This means that each target patient for whom the PAI

prediction is estimated is excluded from the model to avoid overfitting (DeRubeis, Cohen, et al., 2014; Efron & Gong, 1983).

Since the introduction of the PAI, it has been applied to multiple data sets comparing different therapies. One study compared CT to IPT for patients with depression and found a PAI value of 8.9 BDI points. This means that patients who had received their optimal treatment had a BDI-II post-treatment score that was 8.9 points lower than if they had received the treatment suboptimal for them (Huibers et al., 2015). More recently, van Bronswijk and colleagues analyzed data from an RCT that compared CT with IPT for patients with depression as well and found a PAI of 5 points on the BDI-II. Interestingly, they compared a long-term PAI with the "regular" post-treatment PAI and found a weak correlation between the two meaning that the PAIs did not show consistency over time (van Bronswijk et al., 2019). When CBT was compared to psychodynamic therapy for patients with depression, a PAI of 1.6 points on the Hamilton-Depression-Rating Scale was found (Cohen et al., 2019). Moreover, the PAI has also been applied to data sets that compare two different treatments for patients with posttraumatic stress disorder (Deisenhofer et al., 2018; Keefe et al., 2018).

Moreover, different variations of the PAI approach have been developed and tested that are aimed at differentiating placebo and antidepressant responders (Webb et al., 2019) and minimizing dropouts (Zilcha-Mano et al., 2016). A problem that comes with the heterogeneity of statistical approaches applied to variable and treatment selection is the lack of coherence in the variables found to predict treatment response in depression (Cohen & DeRubeis, 2018). The lack of replicability can be partially explained by the use of different variable selection methods with the same data set. They may lead to different conclusions about the predictors found to be important (Bleich, Kapelner, George, & Jensen, 2014) which in turn may lead to other treatment recommendations (Cohen et al., 2019). Furthermore, the samples used for

multivariable analyses are often statistically underpowered therefore spurious findings might be treated as clinically informative (Delgadillo & Gonzalez Salas Duhne, 2019; Luedtke, Sadikova, & Kessler, 2019).

1.5 Prescriptive and prognostic predictors

The relationship between a pre-treatment variable and the treatment outcome can be either a *prognostic* or a *prescriptive* one. A *prognostic* variable predicts treatment outcome irrespective of treatment condition. If the treatment response is predicted for only one treatment, only prognostic predictors can be found (Cohen & DeRubeis, 2018). In contrast, a *prescriptive* variable predicts a differential treatment response to two or more treatment modalities (Fournier et al., 2009). The latter is commonly called a moderator (Kraemer, Wilson, Fairburn, & Agras, 2002) and affects the strength or direction of differences in outcome (Baron & Kenny, 1986). When comparing two or more treatments, the *prescriptive* variables predict whether a patient benefits more from one treatment than another. However, the meaning of the term *prescriptive* is not just about making predictions, but about giving a clear treatment recommendation like Beutler and Harwood proposed when they introduced *Prescriptive Psychotherapy* (Beutler & Harwood, 2000). It is important to note that a variable found to be *prognostic* in one study can predict a differential treatment response, thus be *prescriptive* in another (Cohen & DeRubeis, 2018).

There is little consistency in the literature regarding the variables that predict treatment outcome for patients with depression. Fournier and colleagues (2009) found the variables marriage, unemployment, and having experienced a greater number of recent life events to predict superior responses to CT compared to ADM. Another study found somatic complaints, paranoid symptoms, interpersonal self-sacrificing, attributional style focused on achievement goals, and the number of life events in the past year to predict a favorable

treatment outcome in CT compared to IPT (Huibers et al., 2015). Moreover, DeRubeis and colleagues (2014) found the absence of a comorbid personality disorder, marriage, employment, a greater number of stressful life events, and a greater number of prior antidepressant trials to predict favorable outcome in CBT compared to ADM. These examples show the diversity of variables that are being evaluated and the heterogeneity in the found results. Consequently, the use of these predictors for informed treatment selection becomes a difficult task for clinicians.

1.6 Reference to the two studies presented

The two studies of this doctoral thesis use the methodology of the PAI to advance the field of treatment selection in clinical psychology. Although the methodology is the same, the two studies provide a new contribution to the field of treatment selection for patients with depression as they both investigate treatment modalities that have not been evaluated yet in this sense. **Study 1** compares an assimilative integration to a standard psychotherapy whereas **Study 2** compares blended internet- and face-to-face treatment to routine care. Chapter 2 presents sample, the data analytical strategy and the most important results of the two studies. Chapter 3 presents the limitations of the studies and problems in the field of treatment selection. Moreover, it discusses the future of treatment selection and personalizing treatments. Furthermore, the results of the two studies are put into the context of process outcome research and an example of the implementation of treatment selection in clinical practice is given.

2 Summary of the studies

The following chapter presents the two studies of this doctoral thesis in a summarized form:

Study 1:

Friedl, N., Berger, T., Krieger, T., Caspar, F., & Grosse Holtforth, M. (2019). Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE). *Psychotherapy Research*, 1-13.

Study 2:

Friedl, N., Krieger, T., Chevreul, K., Hazo, JB., Holtzmann, J., Hoogendoorn, M., Kleiboer, A., Mathiasen, K., Urech, A., Riper, H., & Berger, T., (in press). Using the Personalized Advantage Index for individual treatment allocation to blended treatment or treatment as usual for depression in secondary care.

To provide a good overview, the studies are each divided into four sections and briefly summarized below: 1) goals and research questions of the study, (2) sample of the study, (3) data analytical strategy and (4) results.

2.1 Study 1: Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE).

2.1.1 Goal of the study and research questions

Chapter 1 referred to the need for personalized treatments for patients with depression because of the individual differences in the course of illness and the high variability in patient's treatment response. Multiple studies have used the PAI approach to solve this issue but up to date, only two standard treatments have been compared. That is why Study 1 compared CBT to an assimilative integration that combines CBT with exposure and emotion-focused elements (CBT-EE). The most important predictors of treatment outcome in both conditions were identified and the question of whether model-determined treatment allocation using predictive information results in better treatment outcomes was addressed.

2.1.2 Sample of Study 1

Study 1 used the data of an RCT that compared the efficacy of CBT to CBT-EE for patients with depression (Grosse Holtforth et al., 2019). The study was conducted in the psychotherapy outpatient clinic of the University of Zurich's Department of Psychology. Participants had to meet the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000) criteria for MDD; be 18-65 years old and give informed consent. The following exclusion criteria were applicable: (1) current or lifetime psychotic disorder, (2) bipolar disorder (current or lifetime), (3) schizotypic, antisocial or borderline personality disorder, (4) current substance dependence, (5) acute suicidality, (6) being in psychological treatment for depression, and (7) health conditions that require medications that potentially exacerbate depression. Patients were randomized to receive 22 sessions of either CBT or CBT-EE. Therapists provided both treatments in order to minimize the risk of

confounding treatment effects with therapist effects. Of the 149 patients included, 72 were randomized to CBT and 77 to CBT-EE. Patient characteristics did not differ significantly in both treatment groups. The majority of patients was female (53.9% and 59.2%), single or married and the length of the current depressive episode was shorter than two years. The majority of patients did not have a comorbid anxiety disorder or personality disorder and did not take antidepressant medication.

2.1.3 Data analytical strategy of Study 1

The primary outcome measure for Study 1 was the Beck Depression Inventory-II (BDI-II) (Hautzinger, Keller, & Kühner, 2006). Cronbach's alpha was .88 in the current sample. 42 potential predictors were included in the analysis that can be classified into three categories: (1) sociodemographic variables, (2) depression related variables and (3) other baseline measures. Several of the other baseline measures target constructs that are particularly relevant to CBT-EE. The questionnaires that were used to measure them are the German version of the Ambivalence over the Expression of Emotion Scale (AVEX; Deighton & Traue, 2006), the German Questionnaire "Selbsteinschätzung Emotionaler Kompetenzen" (SEK-27; Berking & Znoj, 2008), the short German version of the Generalized Expectancies for Negative Mood Regulation Scale (NMR; Backenstrass, Pfeiffer, Schwarz, Catanzaro, & Mearns, 2008) and the subscales "mindfulness", "over-identification", "isolation", "self-kindness", "self-judgment" and "common humanity" of the German version of the Self-Compassion Scale (SCS; Hupfeld & Ruffieux, 2011). Furthermore, motive importance was measured with the Inventory of Approach and Avoidance Motives (IAAM; Grosse Holtforth & Grawe, 2000), the satisfaction of the motives with the Incongruence Questionnaire (INC; Grosse Holtforth & Grawe, 2003) and the cognitive and behavioral avoidance measured with the German version of the Cognitive-Behavioral Avoidance Scale (CBAS; Röthlin et al., 2010). Constructs that were thought to be similarly relevant to CBT and CBT-EE were

comorbidities, the general symptomatology assessed with the SCL-K-9 (Klaghofer & Brähler, 2001), basic dimensions of interpersonal problems ("love" and "dominance") measured with the short version of the Inventory of Interpersonal Problems (IIP-32; Thomas, Brähler, & Strauß, 2011), dysfunctional attitudes measured by the subscales "recognition by others" and "performance evaluation" of the Dysfunctional Attitudes Scale (DAS; Hautzinger, Joormann, & Keller, 2005), rumination measured with the subscales "self-focused rumination" and "symptom focused rumination" of the German version of the Ruminative Response Scale-short form (RRS; Kühner, Huffziger, & Nolen-Hoeksema, 2007), self-esteem assessed with the Rosenberg Self-Esteem Scale (RSES; Roth, Decker, Herzberg, & Brähler, 2008), strength of resources was measured with the Bernese Resource Inventory (RES; Trösken & Grawe, 2002), perceived self-efficacy was assessed with the Self-Efficacy Scale (SWE; Jerusalem & Schwarzer, 1999), therapy expectations assessed by the subscales "hope for improvement" and "fear of change" of the Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV; Schulte, 2005), and finally quality of life/well-being was measured with the WHO-QOL (World Health Organization, 1995) and the WHO-5 questionnaires (Brähler, Mühlan, Albani, & Schmidt, 2007) were used as well.

After removing missing baseline data with the R package *missForest* (Stekhoven, 2013), the data set was split into two subsets, the CBT condition, and the CBT-EE condition. Then, BMA (Fragoso, Bertoli, & Louzada, 2018) was used for variable selection with treatment outcome as the dependent variable and to compute separate linear regression models for each treatment condition. Posterior probabilities were used to evaluate the relative importance of each potential predictor. Variables with a posterior probability over 0.5 are defined as important because they are included in over 50% of all the models. In total, 30,000 linear regression models were estimated. For generating the PAI, regression models using a leave-one-out approach (Efron & Gong, 1983) were estimated for each patient. This means

that each patient for whom the PAI prediction is estimated is excluded from the model to avoid overfitting. For each patient, a *factual prediction* (post-treatment BDI-II score of the treatment the patient has received) and a *counterfactual prediction* (post-treatment BDI-II score of the intervention the patient did not receive) were estimated. Finally, the observed change scores are compared. The PAI is the size of the predicted difference of receiving the treatment with the greater predicted benefit.

2.1.4 Results of Study 1

Variables predicting outcome in CBT-EE

The most important predictors found in the CBT-EE condition were pre-treatment BDI-II score (Prob = 100%), age (Prob = 99.9%), comorbid axis-II disorder (Prob = 97.9%), being separated or divorced (Prob = 97.6), self-focused rumination (Prob = 97.3%), the SCL score (Prob = 96.6%), comorbid anxiety (Prob = 93.7%), hope for improvement (Prob = 91.1%) and having accomplished an apprenticeship (Prob = 80.6%). A higher pre-treatment score, higher age, a comorbid axis-II disorder, being separated or divorced, higher self-focused rumination, a higher SCL score, comorbid anxiety and having accomplished an apprenticeship predicted a higher post-treatment score. Higher hope for improvement predicted lower post-treatment scores.

Variables predicting outcome in CBT

Based on the posterior probabilities, the most important variables predicting treatment outcome in CBT were pre-treatment BDI-II score (Prob = 100%), recurrent depression (Prob = 83.3%), the number of previous depressive episodes (Prob = 81.1%), the subscale avoidance in the incongruence questionnaire (Prob = 67.6%) and gender (Prob = 50.8%). A higher pre-treatment score, recurrent depression, a higher number of episodes, higher avoidance and male gender predicted higher posttreatment scores.

The PAI

The true error of the BDI-II post-treatment score predictions was 6.74. This stands for the average absolute difference between the predicted and actual scores across all patients. After treatment, patients who received their optimal treatment had a mean BDI-II score of 8.65 (SD=7.49, n=62) whereas patients who were classified as having received their suboptimal treatment had a mean BDI-II score of 10 (SD=8.75, n=61). This results in an average PAI of 1.35 BDI-II points which can be read as follows: if patients had received their model-determined optimal treatment, their post-treatment BDI-II score would have been more than 1 point lower than if they had received the suboptimal treatment. For 46% of the patients the PAI was 5 or greater which means that for these patients, a substantial difference was predicted between the two treatments. This reference point was defined and used in prior studies as well (Hiroe et al., 2005; Huibers et al., 2015).

2.1.5 Manuscript of Study 1

Friedl, N., Berger, T., Krieger, T., Caspar, F., & Grosse Holtforth, M. (2019). Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE). *Psychotherapy Research*, 1-13.

Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE)

Nadine Friedl [1], Thomas Berger [1], Tobias Krieger [1], Franz Caspar [1]
& Martin Grosse Holtforth [1,2]

[1] Department of Clinical Psychology and Psychotherapy, University of Bern, Switzerland
[2] Division of Psychosomatic Medicine, Department of Neurology, Inselspital, Bern University Hospital, Bern, Switzerland

*Address for correspondence
Department of Clinical Psychology and Psychotherapy, University of Bern
Fabrikstrasse 8, CH-3012 Bern, Switzerland
Tel: +41 (0) 31 631 36 91
E-mail: nadine.friedl@psy.unibe.ch

Abstract

Even though different psychotherapeutic interventions for depression have shown to be effective, patients suffering from depression vary substantially in their treatment response. The goal of this study was to answer the following research questions: (1) What are the most important predictors determining optimal treatment allocation to cognitive behavioral therapy (CBT) or CBT with integrated exposure and emotion-focused elements (CBT-EE)?, and (2) Would model-determined treatment allocation using this predictive information result in better treatment outcomes? Bayesian Model Averaging (BMA) was applied to the data of a randomized controlled trial comparing the efficacy of CBT and CBT-EE in depressive outpatients. Predictions were made for every patient for both treatment conditions and an optimal versus a suboptimal treatment was identified in each case. An index comparing the two estimates, the Personalized Advantage Index (PAI), was calculated. Different predictors were found for both conditions. A PAI of 1.35 BDI-II points for the two conditions was found and 46% of the sample was predicted to have a clinically meaningful advantage in one of the therapies. Although the utility of the PAI approach must be further confirmed in prospective research, the present study study promotes the identification of specific interventions favorable for specific patients.

Keywords: Personalized Advantage Index, treatment selection, depression, psychotherapy, CBT, precision medicine

Clinical or Methodological Significance of this Article: Depression is a heterogenous disorder and patients differ based on numerous baseline characteristics. Some patients benefit more from one psychological treatment than from the other. As multiple treatments with empirically supported efficacy are available, it is difficult for clinicians as well as for patients to know which treatment to choose. Methodological advances could enable clinicians to make a better treatment selection based on empirically identified predictors of post-treatment

outcome. This study's results show that by using BMA, it is possible to make treatment outcome predictions based on a limited set of baseline variables. The PAI findings show that for most patients it does not seem to play a major role if they receive standard CBT or a CBT treatment that assimilatively integrates exposure principles via emotion focused interventions (CBT-EE). However, for 46% of the sample, a substantial difference was predicted for the two treatments.

Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE)

By 2030, depression will be one of the leading causes of the global burden of disease (Mathers & Loncar, 2006). Nowadays, it is already one of the most common problems seen in clinical practice and is associated with great suffering of the affected patients as well as high societal costs (World Health Organization, 2008). Different effective psychotherapeutic approaches are available for the treatment of depression. Two of the most well-studied and frequently applied psychotherapies for major depressive disorder (MDD) are cognitive behavior therapy (CBT) and interpersonal psychotherapy (IPT) (Barth et al., 2016; Cuijpers et al., 2013; Lemmens et al., 2018). The question of whether all psychotherapies are equivalent has been addressed early on by Luborsky, Singer and Luborsky (1975) and a bit later by Stiles, Shapiro and Elliott (1986). They discuss the paradox of no differential effectiveness of different psychotherapies despite their technical diversity. Multiple meta-analyses suggest that the major psychotherapeutic approaches for depression do not differ with regard to their efficacy (Cuijpers & Dekker, 2005; Wampold, Minami, Baskin, & Tierney, 2002).

However, whereas the average effects of different psychotherapeutic approaches do not differ on a group level, patients with depression vary substantially with regard to their individual treatment response and their individual illness course (Cuijpers et al., 2010). Increasing evidence suggests that patient characteristics and traits moderate the efficacy of different treatments at an individual level (Fournier et al., 2009), so that differential responses to treatments are not a scarcity (Simon & Perlis, 2010). Thus, by finding for each individual the psychological treatment that works best for him or her, it may be possible to improve the overall treatment response rates (Hollon, Cohen, Singla, & Andrews, 2019).

In this sense, the approach of precision medicine, which has a long tradition in medicine, tries to tailor medical treatments to the specific characteristics and needs of the patients (Deisenhofer et al., 2018; Hamburg & Collins, 2010). Accordingly, the use of prediction algorithms to improve and support clinical decision-making becomes increasingly common in the medical field. In clinical psychology and psychotherapy, the use of algorithms to predict the optimal treatment for an individual is growing as well (van Bronswijk et al., 2018). One example is the "Personalized Advantage Index" approach which has the goal of identifying patients with a certain disorder (e.g., MDD) for whom one treatment works better than another (Cohen & DeRubeis, 2018). If there are two comparably effective treatments to choose from, the Personalized Advantage Index (PAI) identifies the treatment that predicts a better treatment outcome for a given patient with particular characteristics and provides a quantitative estimate of how much a particular treatment is better for this patient than another (DeRubeis et al., 2014). By this, the PAI approach promises to increase the likelihood of improving a patient's mental health by identifying subsets of a given patient population who will likely show a certain response to a given treatment (Hollon et al., 2019). So far, its feasibility and relevance have been shown in several studies on the treatment of depression (Cohen et al., 2017; DeRubeis et al., 2005; Huibers et al., 2015). Huibers and colleagues (2015) for example, compared Cognitive Therapy (CT) with IPT and found that the patients randomized to their predicted optimal treatment (either CT or IPT) had an observed mean end score on the Beck Depression Inventory (BDI) of 11.8, while those who received their predicted non-optimal treatment had an end-BDI of 17.8. Similarly, Cohen and colleagues (2017) compared CBT with psychodynamic therapy (PDT) and found an average post-treatment Hamilton Depression Rating Scale (HDRS) score of 1.6 points lower for the patients who received their indicated treatment compared to non-indicated treatment. These

results indicate that being able to identify the best treatment for an individual with depression may make health care delivery more efficient (Hollon et al., 2019).

So far, research on the PAI has focused on comparing standard manualized psychotherapeutic approaches, such as CBT with PDT. What has not been done so far is the comparison of a standard psychotherapy with a form of psychotherapy that integrates additional interventions into the standard approach following the principles of assimilative integration. Assimilative integration entails the systematic inclusion of techniques and attitudes of a supplementary therapy into a therapist's primary therapeutic approach (Castonguay, Newman, Borkovec, grosse Holtforth, & Maramba, 2005; Messer, 1992). The difference to prior studies is that the treatments being compared are largely similar with the exception that the "new" treatment integrates additional interventions and principles that the standard treatment did not while preserving the overall therapeutic approach.

In the current study, Exposure-Based Cognitive Therapy (EBCT-R; grosse Holtforth et al., 2012; Hayes, Beevers, Feldman, Laurenceau, & Perlman, 2005) is compared with standard CBT. EBCT-R integrates principles of exposure therapy operationalized via interventions from emotion-focused therapy (e. g., empty chair and two-chair interventions) into a cognitive-behavioral treatment in an attempt to increase a patient's level of emotional processing. Emotional processing is seen as a key mechanism of change for the improvement in depression (Grosse Holtforth et al., 2019; Hayes, 2015; Pinheiro, Mendes, Silva, Gonçalves, & Salgado, 2018). Even though the application of exposure principles in depression treatment is fairly new, EBCT-R has been associated with significant reductions in depressive symptoms and has shown to be equally effective as CBT and produce effect sizes of similar size (grosse Holtforth et al., 2012; Grosse Holtforth et al., 2019; Hayes et al., 2005). At this point, it is important to note that the original trial (Grosse Holtforth et al., 2019) did not intend to invent a new therapeutic approach, but rather empirically test the inclusion of

additional elements into an existing approach according to the principles of assimilative integration. In order to stress both, the underlying standard approach (CBT) as well as the integrated elements (i. e., use of emotion-focused interventions as a way to realize exposure principles), we decided to use the abbreviation “CBT-EE” instead of the previously used “EBCT-R” for this integrative condition. Consequently, the design of the trial as well as the use of the PAI approach promise to improve treatment allocation by empirically identifying individual outcome predictors and providing a model for optimal treatment allocation.

To reach this goal, the present study set out to answer the following research questions: (1) What are the most important predictors determining optimal treatment allocation to CBT or CBT-EE? and (2) Would model-determined treatment allocation using this predictive information result in better treatment outcomes?

Method

This study was based on a randomized controlled trial (RCT) comparing the efficacy of CBT and CBT-EE (Grosse Holtforth et al., 2019). The study was conducted at the outpatient clinic of the Department of Psychology at the University of Zurich. Each patient was randomized to receive 22 sessions of either CBT or CBT-EE. In order to reduce the risk of confounding treatment effects with therapist effects, therapists provided both treatments and were assigned an equal number of patients in both conditions. Therapists in the CBT condition were instructed to refrain from using emotion-focused interventions. Of the 149 individuals that were included, 72 were randomized to CBT and 77 were randomized to CBT-EE. Participants met the following inclusion criteria: (1) meeting Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric American Psychiatric Association, 2000) criteria for MDD, (2) being 18 to 65 years old and (3) giving informed consent. Details about the study design and a sample description are described elsewhere (grosse Holtforth et al., 2019).

Measures

Primary outcome. The primary outcome measure for the present study was the Beck Depression Inventory-II (BDI-II; Hautzinger, Keller, & Kühner, 2006). The BDI-II is the most common used self-assessment questionnaire to evaluate the severity of depression. The reliability, validity and sensitivity to change of the German BDI-II has been shown to be satisfactory in prior studies (Kühner, Bürger, Keller, & Hautzinger, 2007). In the current sample, Cronbach's alpha was .88.

Predictor variables. Following an exploratory approach, a total of 42 potential predictors were included in the analysis. Various variables that had been measured at baseline were included as possible post-treatment predictors. To provide a better overview of the variables included in the analysis, they can be classified into three categories: (1) sociodemographic variables, (2) depression related variables and (3) other baseline variables. In the RCT, some of the other baseline variables had been chosen theoretically to assess constructs particularly relevant to CBT-EE. These variables are: ambivalence over the expression of emotion measured using the German version of the Ambivalence over the Expression of Emotion Scale (AVEX; Deighton & Traue, 2006), emotion regulation skills measured with the German questionnaire "Selbsteinschätzung Emotionaler Kompetenzen" (SEK-27; Berking & Znoj, 2008), negative mood regulation measured with the short German version of the Generalized Expectancies for Negative Mood Regulation Scale (NMR; Backenstrass, Pfeiffer, Schwarz, Catanzaro, & Mearns, 2008), as well as self-compassion measured using the subscales "mindfulness", "over-identification", isolation", "self-kindness", "self-judgement" and "common humanity" of the German version of the Self-Compassion Scale (SCS; Hupfeld & Ruffieux, 2011). For some of the other baseline variables, the assumed relevance to CBT-EE varied between the subscales, i.e., motive importance measured with the Inventory of Approach and Avoidance Motives (IAAM; grosse

Holtforth & Grawe, 2000), satisfaction of motives measured with the Incongruence Questionnaire (INC; grosse Holtforth & Grawe, 2003), and the cognitive and behavioral avoidance measured with the German version of the Cognitive-Behavioral Avoidance Scale (CBAS; Röthlin et al., 2009). The remaining baseline variables were assumed to be similarly relevant to CBT and CBT-EE, i.e., comorbidities (anxiety, axis-I, axis-II), general symptomatology assessed by the SCL-K-9 (Klaghofer & Brähler, 2001), basic dimensions of interpersonal problems ("love" and "dominance") assessed with the short version of the Inventory of Interpersonal Problems (IIP-32; Thomas, Brähler, & Strauß, 2011), as well as dysfunctional attitudes measured by the subscales "recognition by others" and "performance evaluation" of the Dysfunctional Attitudes Scale (DAS; Hautzinger, Joormann, & Keller, 2005), rumination measured with the subscales "self-focused rumination" and "symptom focused rumination" of the German version of the Ruminative Response Scale-short form (RRS; Kühner, Huffziger, & Nolen-Hoeksema, 2007) self-esteem was measured using the Rosenberg Self-Esteem Scale (RSES; Roth, Decker, Herzberg, & Brähler, 2008), strength of resources measured with the Bernese Resource Inventory (RES; Trösken & Grawe, 2002), perceived self-efficacy measured with the Self-Efficacy Scale (SWE; Jerusalem & Schwarzer, 1999), therapy expectations assessed by the subscales "hope for improvement" and "fear of change" of the Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV; Schulte, 2005), as well as quality of life/well-being measured by the WHO-QOL (The World Health Organization quality of life assessment, 1995) and the WHO-5 questionnaires (Brähler, Mühlan, Albani, & Schmidt, 2007).

Data analytical strategy

Generally, we followed a bottom-up analytical approach with regard to predictor variables. This means that we treat the different types of predictors equally in the data analysis, regardless of their assumed particular relevance to CBT-EE.

Missing data and removing linear combinations. As in previous work using the PAI, we focused on those participants for which end of treatment BDI-II scores were available (DeRubeis et al., 2014a; Huibers et al., 2015). This leaves us with $N = 123$, representing 83% of the total sample. Distributed over the two conditions, there were 63 patients in the CBT condition and 60 patients in the CBT-EE condition. Missing data in the baseline measures was imputed using the R package *missForest* (Stekhoven, 2013), which uses a random forest approach trained on the observed values of the data to predict missing values. Missing values were found in the variables: marital status, education, chronic depression, number of episodes, the WHO-5, AVEX, NMR and SCL. This method treats the missing value problem as a prediction problem (Tang & Ishwaran, 2017). The number of episodes had 13 missings (10.6%), marital status had 7 missings (5.7%), education, AVEX and NMR had four missings (3.3%) and chronic depression, WHO-5 and the SCL had two missings (1.6%). It is especially useful when dealing with mixed-type data, as is the case in the present study because it can handle categorical and continuous variables simultaneously (Stekhoven & Buhlmann, 2012). Because of its small imputation error, *missForest* has been shown to outperform other imputation methods and to be highly accurate (Waljee et al., 2013). All further analyses are based on the imputed data set. The predictors included in the analyses were centered whereby continuous variables were centered at the grand mean and dichotomous variables were dummy coded to -0.5 and 0.5 (Kraemer & Blasey, 2004).

Before building the regression model, we checked for linear combinations of features as an indication of redundancy because this may cause problems in linear regression. Although exact linear combinations are rare, they can occur if there is a larger number of features (Forte, 2015, pp. 27-28). We used the package *caret* because its *findLinearCombos* function can identify exact linear combinations of features (Kuhn, 2018). For the current data set, results indicated that no exact linear combination of other features are to be found.

Bayesian Model Averaging (BMA). A standard data analysis approach to identify baseline variables that predict outcome in one treatment versus another would rely on one single model that can only include a limited number of potential predictors. A commonly used rule of thumb states that at least 10 observations per predictor are necessary to not exceed an acceptable level of bias (Peduzzi, Concato, Feinstein, & Holford, 1995; Vittinghoff & McCulloch, 2007). However, only a small number of predictors could have been included if we followed this approach. Furthermore, relying upon a single selected model may lead to overconfidence in the conclusions drawn regarding quantified associations. This might happen because the uncertainty in the model is ignored and because an alternative model with a different subset of predictors may fit the data equally well (Raftery, Madigan, & Hoeting, 1997; Stock et al., 2014). The problem with overfitting is that the models might fail to replicate in future samples, thus creating considerable uncertainty about the scientific value of the findings (Babyak, 2004). A possible remedy for this dilemma that might provide a better predictive ability is to account for model uncertainty through Bayesian Model Averaging (BMA). The BMA method applies Bayesian inference for model selection and prediction resulting in less risky predictions and simpler model choice criteria (Fragoso, Bertoli, & Louzada, 2018). The Bayes factor is a way to investigate the evidence in favor of a null hypothesis, thus it is a method to quantify the evidence that supports a scientific theory (Kass & Raftery, 1995). BMA incorporates model uncertainty into the parameter estimates and inferences by estimating a posterior probability that each considered model is the correct one, given the data. Inferences can then be based on posterior means (weighted averages) of quantities across models (Raftery, 1995). To be precise, BMA averages over all possible sets of predictors and makes it possible to choose the most probable one. In summary, BMA is a way to derive sharper predictions from the data particularly in cases with many possible regressors but limited observations. BMA's predictive performance has supported empirically

in several studies (Fernandez, Ley, & Steel, 2001a, 2001b; Ley & Steel, 2009; Steel, 2011; Stock et al., 2014).

The data frame was split into two subsets, the CBT condition and the CBT-EE condition. Then, separate linear regression models were computed for each treatment condition (Deisenhofer et al., 2018) using the R package *BAS* (Clyde, 2018). We performed a Bayesian adaptive sampling (BAS) without replacement for variable selection in linear models using the function *bas.lm* with treatment outcome (BDI-II post-treatment score) as dependent variable. As mentioned before, posterior probabilities for each potential predictor are calculated and used to evaluate the relative importance of each variable. Posterior probabilities were also used as the criterion for evaluating the model performance. The following considerations determined choosing the appropriate *bas.lm* function: we used the *"ZS-null"* criterion for the priors, which is a Laplace approximation to the *Jeffreys-Zellner-Siow* (JZS) prior for the integration of alpha = 1 only. The *JZS prior* uses the Jeffreys prior on sigma and the Zellner-Siow Cauchy prior on the coefficients. The optional parameter 'alpha' can be used to control the squared scale of the prior, where the default is alpha=1 (Clyde, 2018). We decided to use *"MCMC+BAS" method*, which runs an initial Markov chain Monte Carlo (MCMC) algorithm to calculate marginal inclusion probabilities and then samples without replacement as in BAS. The combined option is recommended as it provides estimates with low bias compared to the sampling of BAS alone (Clyde, Ghosh, & Littman, 2011). The number of models was set to 30,000 because it was assumed that each additional model will add only a small increment to the cumulative probability. This means that it leads to no nameworthy differences in posterior distributions. This assumption was later examined and analyzed.

Personalized Advantage Index (PAI). The first step in generating the PAI is to estimate regression models using a leave-one-out approach, also known as jackknife

(DeRubeis et al., 2014; Efron & Gong, 1983). In this procedure, each target patient for whom the PAI prediction is estimated, is excluded from the model to avoid overfitting. For each of the 123 patients, two predictions were estimated: the post-treatment BDI-II score of the treatment the patient has received (factual prediction) and a post-treatment BDI-II score of the intervention the patient did not receive (counterfactual prediction). For the estimation of the factual prediction, the values of all the available pre-treatment variables, including their treatment assignment, were entered in the regression model. In the next step, using the predicted scores, we looked into the observed change scores of each patient which is the size of the predicted difference of receiving the treatment with the greater predicted benefit (optimal treatment) versus the non-optimal treatment. This provides us with a quantitative estimate, a predicted advantage, namely the PAI (DeRubeis et al., 2014). Higher absolute values of the PAI stand for stronger predicted benefits of one treatment over another. Up to today, PAI values vary from 1.6 on the Hamilton-Depression-Rating Scale (Cohen et al., 2019) up to 8.9 on the BDI-II (Huibers et al., 2015).

Results

The importance of the potential predictors was evaluated based on their marginal posterior inclusion probabilities. Values above 0.5 indicate that the predictor has been included in more than half of the models, thus in over 15,000 models in the present case. In the following, we first report on the best five models for each treatment condition, and then we report the resulting PAI for the whole sample.

Variables predicting outcome in CBT-EE

The five best models predicting depression severity at post-treatment in the CBT-EE condition are shown in Table 1. For each model, the Bayes Factor, number of predictors, R^2, log marginal likelihood and the posterior probabilities are provided. The first model (Model 1) seemed to fit the data best because it had the largest Bayes Factor and largest posterior

probability (0.03). Consequently, model 1 was selected as our final predictive model of post-BDI-II score in the CBT-EE condition.

Table 1

Five best models CBT-EE

Fit indices	Model 1	Model 2	Model 3	Model 4	Model 5
Bayes Factor	1	0.9228315	0.6399822	0.4767033	0.4203144
Number of variables	13	12	12	12	14
R^2	0.7767	0.7614	0.7578	0.7548	0.7823
log marginal likelihood	19.55932	19.4790081	19.1130018	18.8184556	18.6925643
Posterior probabilities	0.0335	0.0309	0.0215	0.016	0.0141

While the selected model includes 13 variables in total, the strongest predictors of the post treatment BDI-II score in the CBT-EE condition included pre-treatment BDI-II score (Prob = 100%), age (Prob = 99.9%), comorbid axis-II disorder (Prob = 97.9%), being separated or divorced (Prob = 97.6%), self-focused rumination (Prob = 97.3%), the SCL score (Prob = 96.6%), comorbid anxiety (Prob = 93.7%), hope for improvement (Prob = 91.1%) and having accomplished an apprenticeship (Prob = 80.6%). A higher pre-treatment score, higher age, a comorbid axis-II disorder, being separated or divorced, higher self-focused rumination, a higher SCL score, comorbid anxiety and having accomplished an apprenticeship predicted a higher post-treatment score. Higher hope for improvement predicted lower pre-treatment scores. The effects of other variables appeared minimal due to their small posterior probabilities. See Table 2 for the posterior probabilities of the 13 variables included in the best model (Model 1). See Appendix 1 in the supplemental online material for the complete list of variables and their inclusion probabilities.

Table 2

BMA results for the 13 variables included in the best model of the CBT-EE condition

Baseline variables	Prob (%)	Posterior mean (SD)
Sociodemographic variables		
age	99.9	0.2818 (0.07252)
separated/divorced	97.6	-7.627 (0.009889)
apprenticeship	80.6	-2.546 (1.857)
Depression related variables		
pre-treatment BDI-II	100	-0.7006 (0.1160)
prior antidepressant trials	43.2	1.162 (1.785)
Other baseline measures		
comorbid axis II	97.9	-5.180 (2.068)
RRS (self-focused rumination)	97.3	-0.5317 (0.2229)
SCL-K-9	96.6	3.822 (1.674)
comorbid anxiety	93.7	3.527 (2.217)
PATHEV (subscale hopfulness)	91.1	-1.936 (1.080)
DAS subscale performance evaluation	42.5	0.04713 (0.06966)
INC (avoidance)	23.6	0.4941 (1.206)
PATHEV (subscale fear)	23.4	0.2480 (0.6715)

Note. BMA= Bayesian Model Averaging; RRS= Ruminative Response Scale; SCL-K-9= Symptom Checklist; PATHEV= Patient Questionnaire on Therapy Expectation and Evaluation; DAS = Dysfunctional Attitudes Scale; INC = Incongruence Questionnaire.

Variables predicting outcome in CBT

Table 3 gives an overview of the five best models to predict treatment outcome for CBT. Based on the posterior probabilities and the Bayes Factor, Model 1 is the best model. For that reason, this model including six variables was selected as the final predictive model for the CBT condition.

Table 3

Five best models for CBT

Fit indices	Model 1	Model 2	Model 3	Model 4	Model 5
Bayes Factor	1	0.2965935	0.2451441	0.2366299	0.2267025
Number of variables	6	7	7	7	6
R^2	0.5197	0.5253	0.5221	0.5215	0.4938
log marginal likelihood	11.7682	10.552804	10.3622878	10.3269389	10.2840802
Posterior probabilities	0.1125	0.0334	0.0276	0.0266	0.0255

Based on the posterior probabilities, the most important predictors for treatment outcome in the CBT condition were pre-treatment BDI-II score (Prob = 100%), recurrent depression (Prob = 83.3%), the number of previous depressive episodes (Prob = 81.1%), the subscale avoidance in the incongruence questionnaire (Prob = 67.6%) and gender (Prob = 50.8%). A higher pre-treatment score, recurrent depression, a higher number of episodes, higher avoidance and male gender predicted higher post-treatment scores. The WHO-QOL reached a posterior probability of 43.9%. The other baseline variables did not seem to contribute to the predictive ability of the model. See Table 4 for the posterior probabilities of the six variables included in the final prediction model. See Appendix 2 in the supplemental online material for the posterior probabilities of all variables measured at baseline.

Table 4

BMA results of the 6 variables included in the best model for CBT

Baseline variables	Prob (%)	Posterior mean (SD)
Sociodemographic variables		
gender	50.8	2.030 (2.524)
Depression related variables		
pre-treatment BDI-II	100	- 0.7702 (0.1444)
recurrent depression	83.3	7.868 (5.390)
depressive episodes	81.1	- 2.461 (1.816)
Other baseline measures		
INC (avoidance)	67.6	2.205 (2.128)
WHO-QOL	43.9	- 1.266 (2.364)

Note. BMA = Bayesian Model Averaging; INC= Incongruence Questionnaire; WHO-QOL = World Health Organisation Quality of Life Questionnaire.

Figure 1 shows the cumulative probability for each of the models in the order they were sampled. It shows that, on the one hand, the cumulative probability is leveling off as each additional model adds only a small increment to the cumulative probability, i.e., that we would have gotten less and less information on the predictive ability of our models by running more models. On the other hand, with a smaller number of models we still gained a lot of additional useful information regarding the cumulative probability. However, as the number

of models approached 30,000 the cumulative probability did not increase very much anymore. Thus, it can be assumed that sampling more than 30,000 models would not lead to any nameworthy differences in posterior distributions and the additional information we would get would level off.

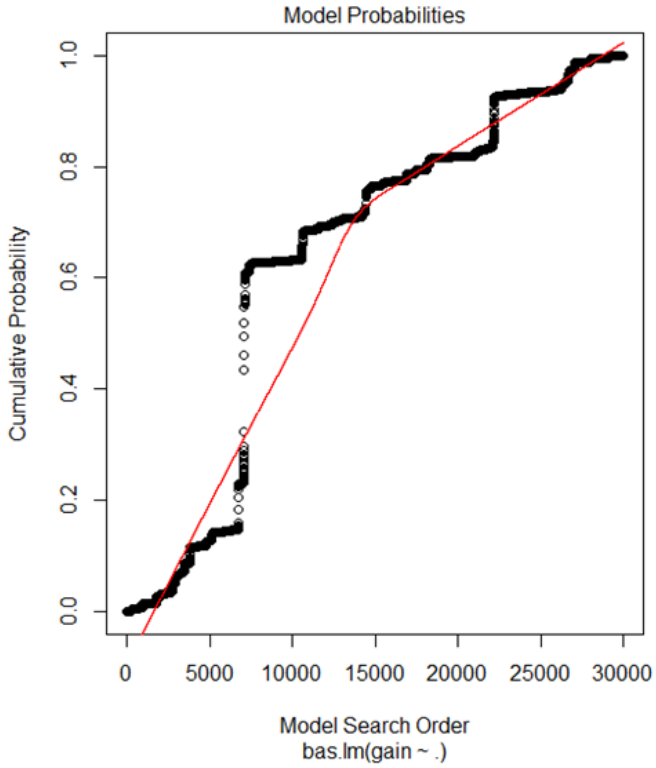

Figure 1. Cumulative probabilities of the CBT model.

Personalized Advantage Index

Using the treatment specific predictors described above, the prediction of a patient's post-treatment BDI-II score was computed separately for each treatment condition. The true error of the BDI-II post-treatment score predictions was 6.74, representing the average absolute difference between the predicted and actual, observed scores across all patients. The true error is to be distinguished from the PAI as it compares the observed outcome of the intervention that the patient received in the RCT with the prediction of the same intervention. Patients who were categorized as having received their optimal treatment had a mean BDI-II post-treatment score of 8.65 (*SD*= 7.49, *n*= 62) whereas patients who were classified as

having received their suboptimal treatment had a mean BDI-II post-treatment score of 10 (*SD*= 8.75, *n*= 61). Figure 2 shows the frequency of predicted BDI-II post-treatment scores for every patient in both the optimal and suboptimal treatment.

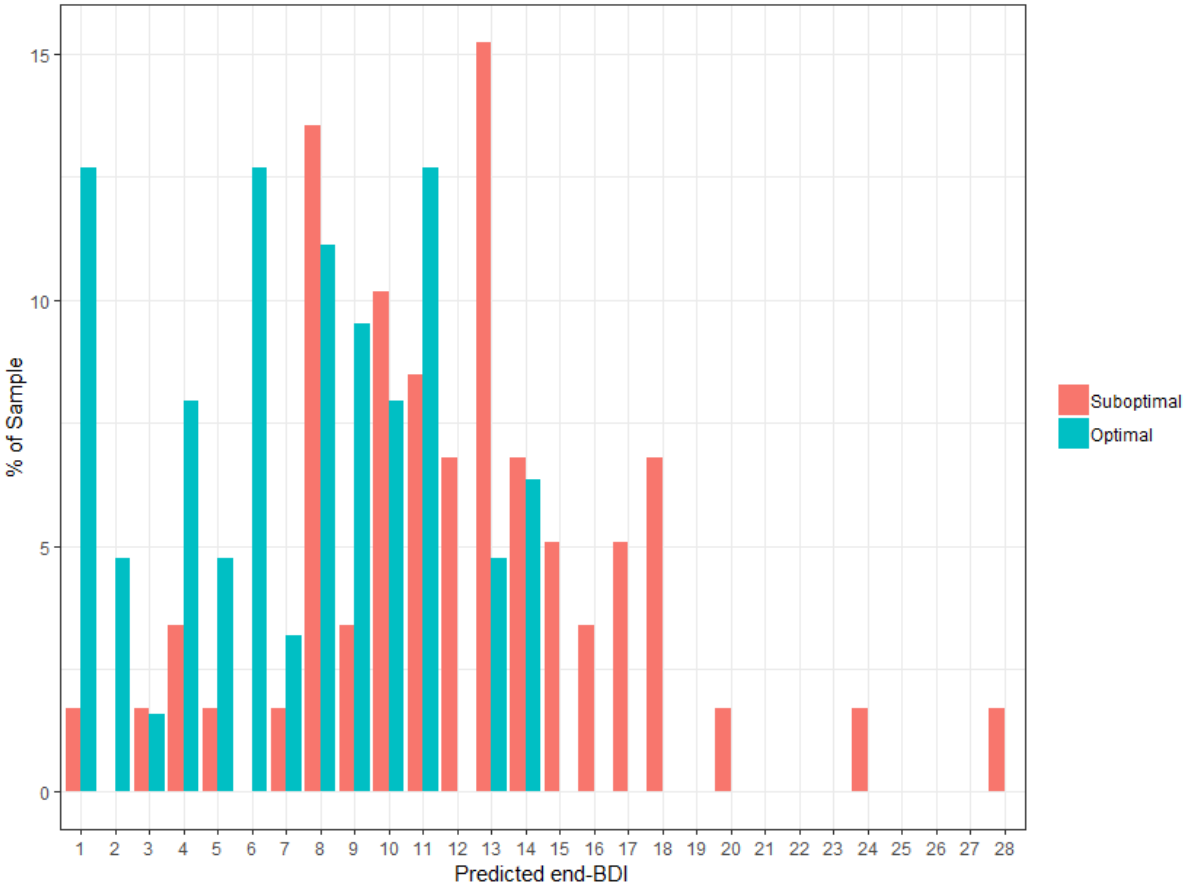

Figure 2. Frequency histogram showing predicted post BDI scores for each patient in their optimal and their suboptimal treatment.

First, an individual PAI was calculated for each patient. Then, an average PAI was calculated as the mean difference in BDI-II scores between the optimal and the suboptimal treatment for each patient, resulting in an average PAI of 1.35. This score can be read as follows: if patients had received their model-determined optimal treatment, their post-treatment BDI-II score would have been more than 1 point lower than if they had received the non-optimal treatment. Figure 3 shows the frequencies of the individual PAIs. For 46% of the patients in this sample the PAI was 5 or greater, indicating that a substantial difference was

predicted between the two treatments. Hiroe and colleagues (2005) have found that a minimal clinically important difference corresponds to a 5-point difference in BDI-II scores. This reference point has been used in prior studies as well (Huibers et al., 2015).

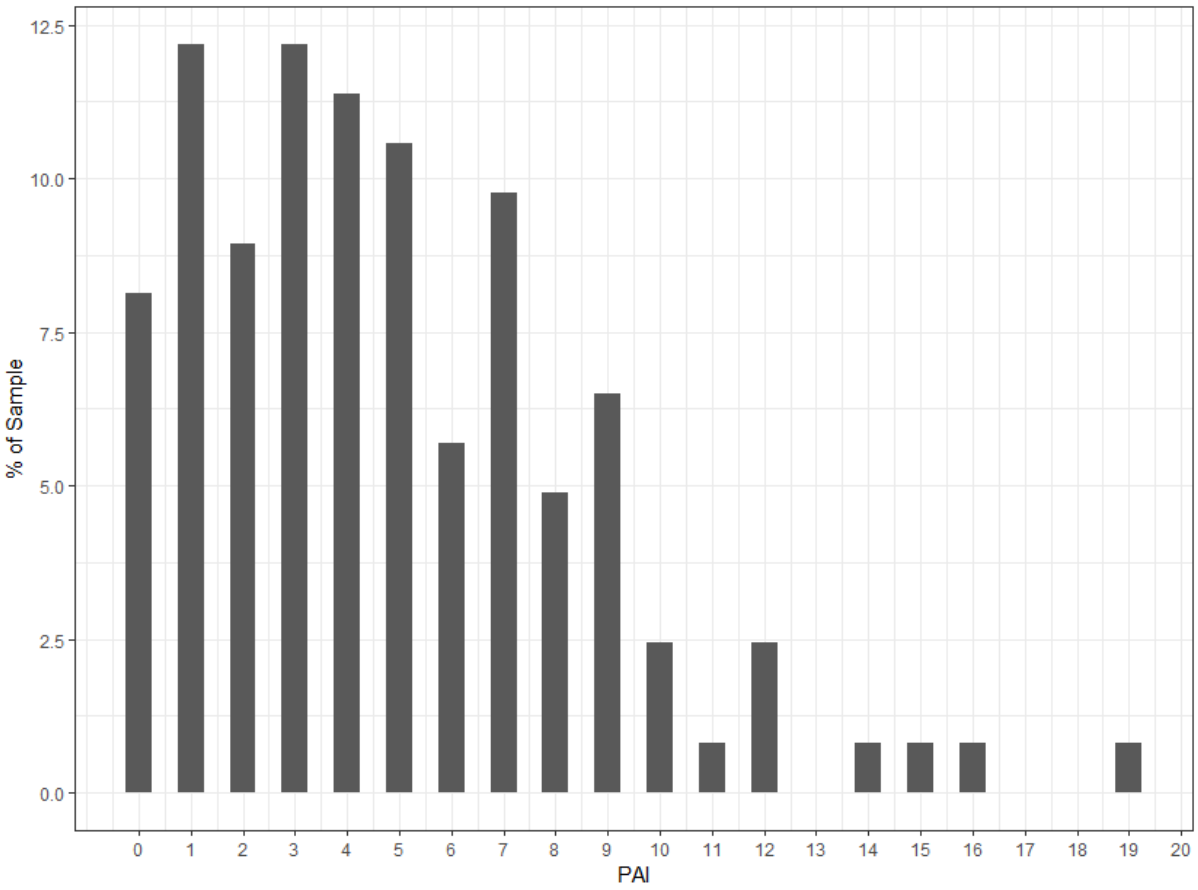

Figure 3. Frequency histogram showing PAI scores.

Discussion

The objectives of the present study were (1) to identify the most important predictors determining optimal treatment allocation to the integrative CBT-EE or standard CBT, and (2) to investigate if model-determined treatment allocation using this predictive information results in better treatment outcomes for an individual patient. With regard to predictors of treatment outcome in each of the interventions, different relevant predictors were identified for CBT-EE and CBT, respectively. In the CBT-EE condition, lower age, not being separated

or divorced, having accomplished a higher education than an apprenticeship, a lower pre-treatment depressive symptomatology, no comorbid anxiety disorder, no comorbid axis-II disorder, a lower psychopathology, lower self-focused rumination, and more hope for improvement predict a better treatment outcome, i.e., a lower post-treatment BDI-II score. In contrast, in the CBT condition, a lower pre-treatment depressive symptomatology, female gender, fewer previous depressive episodes, no recurrent depression and a lower current incongruence regarding avoidance motives predicted a better treatment outcome.

The distinction between prescriptive or prognostic outcome predictors offers an initial interpretation of our findings. Whereas prescriptive variables may support differential indication by predicting whether a patient will benefit more from one treatment in comparison to another, prognostic variable predict treatment outcome regardless of treatment condition. In the present study, the pre-treatment depressive symptomatology is the only *prognostic* predictor, i.e., the only variable that predicts treatment outcome in both conditions. This is in line with many previous studies that found the pre-treatment symptomatology to be a prognostic predictor, in the sense that higher pre-treatment severity is related to worse outcome in all treatment conditions (e.g., Cohen & DeRubeis, 2018; Dinga et al., 2018; van Bronswijk et al., 2018; DeRubeis et al., 2014; Hamilton & Dobson, 2002). Interestingly, none of the baseline variables assumed to be relevant for CBT-EE due to their emotional focus such as emotion regulation skills, negative mood regulation, ambivalence over the expression of emotion, or self-compassion was empirically a relevant outcome predictor in this condition. At least two explanations may account for this finding. Either the therapists responded differentially to the emotional baseline characteristics patients displayed regardless of treatment condition and thereby prevented potential differential predictions (responsiveness; Stiles, Honos-Webb, & Surko, 1998), or the respective emotional pre-treatment characteristic in fact did not have any differential predictive power independent of

how responsive therapists were. Thus, both treatment conditions would be able to deal well with different levels of emotional variables with no advantage of CBT-EE over CBT. As responsiveness may explain any absence of differential or prescriptive predictions, this explanation has to be kept in mind for explaining any absence of prescriptive predictions. Future process research focusing on responsiveness effects in this sample may help to distinguish the effects of responsiveness from non-specificity of predictors.

Regarding further *prescriptive* predictors, our findings partly correspond to previously reported predictors of treatment outcome in CBT. For instance, female gender has been found to predict lower depressive symptomatology at post-treatment in both CBT and IPT (Huibers et al., 2015). Furthermore, previous depressive episodes and recurrent depression negatively predicted outcome in CBT (Blom et al., 2007; Thase, Reynolds, Frank, & Simons, 1994; Fournier et al., 2009). The difference between these previous findings and the results in our study is that in previous studies, female gender and chronicity of depression were prognostic predictors and thus relevant in various conditions. In the current study, these frequently found general predictors of outcome in depression treatments did not predict outcome in the CBT-EE condition. Some of the variables identified as *prescriptive* predictors for CBT-EE in the current study, such as age, education level, and employment have been found to act as either prognostic or prescriptive predictors in CBT, CT or treatment with antidepressant medication (Chekroud et al., 2016; DeRubeis et al., 2014; Fournier et al., 2009; Huibers et al., 2015). Thus, it is rather surprising that we did not find these variables to predict outcome also in the CBT condition.

Interestingly, self-focused rumination and hope for improvement were important predictors in the CBT-EE condition but not in the CBT condition. Apparently, ruminating too much about oneself decreases outcome specifically in CBT-EE. Particularly for patients ruminating too much about themselves, it might be particularly hard to productively engage in

emotion-focused interventions such as two-chair or empty-chair exercises, as a ruminative processing style regarding the self makes it even harder to change maladaptive self-associations as it is intended in emotion-focused interventions (Greenberg, 2002). With regard to the lacking prediction in the CBT condition, this is in line with Huibers and colleagues (2015) who also assessed rumination with the RRS and did not find it to be an important predictor in CBT. A stronger hope for improvement may particularly facilitate process and outcome in CBT-EE, taking into account that emotion-focused interventions maybe particularly challenging and temporarily burdensome for the patient given the inherent activation of past negative emotions. Potentially, having more hope for improvement initially fosters patients' motivation to "endure" burdensome emotions in service of better outcomes in the long run (Westermann, grosse Holtforth, & Michalak, in press).

For the CBT condition, currently making less aversive experiences (lower avoidance incongruence) seems to be associated with better therapy outcomes, or put conversely, making stronger aversive experiences in life at intake decreases the chance for good therapy outcomes. This finding underscores the potential relevance of motivational factors, that have not been examined in earlier research, in the prediction of therapy outcome.

Naturally, the above interpretations regarding prognostic and prescriptive predictors are somewhat speculative, so that further process research is necessary to test these interpretations empirically. In the end, we cannot rule out confounding factors and we cannot make any causal consumptions.

The second goal of the current study was to investigate a model-determined treatment allocation and to calculate the PAI, a measure of the predicted advantage in one therapy compared to the other. Up to date, various forms of treatment have been compared using the PAI (e. g., Cohen et al., 2019; Huibers et al., 2015) but two forms of CBT (one that integrates emotion-focused techniques to foster emotional processing within an exposure framework)

have not been compared using the PAI yet. Growing empirical evidence supports the efficacy and feasibility of depression treatments that integrate principles of change from other orientations (Castonguay, Eubanks, Goldfried, Muran, & Lutz, 2015; Grosse Holtforth et al., 2019). An advantage of treatments based on assimilative integration may be that therapists can stay within the theoretical framework that they are trained in but at the same time benefit from techniques of different approaches that potentially utilize other mechanisms of change (Lampropoulos, 2001). Another advantage of doing PAI research in the field of assimilative integration is that the same therapists can remain within his or her theoretical framework but conduct both of the compared interventions and does not have to be replaced depending on the predicted indication. In addition, the same-therapist design of the original study further facilitates the direct application the results of the PAI studies.

This study's results show that by using BMA, it is possible to make treatment outcome predictions based on a limited set of baseline variables. The present study set out to show that patients might benefit from different treatments despite sharing the same diagnosis. The current study found a PAI of 1.35, indicating that patients could have a post-treatment BDI-II score that is more than one point lower if they receive their model-determined optimal treatment in comparison to a treatment that is not optimal for them. This is a relatively small value compared to other studies that have found PAIs ranging from 1.6 (Cohen et al., 2019) up to 8.9 (Huibers et al., 2015). However, the smaller value could be explained by the fact that previous studies have compared treatments that differ more strongly from each other. In contrast, we have compared two treatments that have many commonalities due to their shared theoretical background in a CBT framework. Furthermore, the same therapists delivered both interventions which could also have lowered the PAI. Given the satisfactory adherence and equal allegiance for both conditions reported in the original study (grosse Holtforth et al., 2019), the PAI findings show that for several patients it does not seem to play a big role if

they receive standard CBT or a CBT treatment that assimilatively integrates exposure principles via emotion focused interventions (CBT-EE). Regardless of the rather small mean PAI, for 46% of the patients in this sample, a substantial difference was predicted between the two treatments as the PAI was 5 or greater. Other studies have found similar percentages ranging from 60% (DeRubeis et al., 2014) to 63% (Deisenhofer et al., 2018; Huibers et al., 2015).

The current study has several limitations. First, the sample size of the current study was relatively small. Due to this, it was not possible to build the models in one subsample and subsequently use another subsample for validation. However, the sample size requirements for an analysis like that could be lower if the assumptions about the underlying population parameters are outside the range of the reported study. Furthermore, studies with smaller samples are justified if they are designed to develop and validate prescriptive prediction scores to be tested further in future hypothesis-driven confirmatory studies (Luedtke et al., 2019). Secondly, it is not known how the model would perform if applied to other populations that have baseline scores outside the range of this population since we do not have a second sample for validation. Future hypothesis-driven confirmatory studies are needed to validate the predictors found in the current study. Moreover, more data of RCTs that compare two different treatments should be made available for this kind of analyses. Particularly, studies that compare a standard treatment to an integrative treatment according to assimilative principles promise to have innovative potential in this context. Finally, another limitation is that therapist effects could not be analyzed in this study. It is known that therapist effects are a robust phenomenon, thus it is possible that they account for the different treatment outcomes in this sample (Johns, Barkham, Kellett, & Saxon, 2019; Wampold & Brown, 2005). The original RCT where the data was drawn from used the “same therapist design”. Even though this design was used in order to minimize therapist effects, it might not have been enough to

remove them completely as allegiance might play a role and because there is not enough information about a potential carry-over effect due to the crossed-therapist design (Grosse Holtforth et al., 2019). Despite these limitations, the current study has the potential to make a significant contribution to the further understanding of depression treatment because it investigates implications for the use of an assimilative integration to CBT for patients with depression. Especially factors found to play a role in the treatment and change process of depression can be considered in clinical practice in order to provide the optimal treatment for each individual. Most importantly, the identified predictors need to be tested prospectively within clinical routine treatment settings in future research. Clinicians should take into account the important predictors found in this study and in previous studies to make an informed treatment recommendation for their patients. Furthermore, this study could help clinicians deciding whether to use their own therapeutic approach or whether it is necessary to include elements from another approach in the form of an assimilative integration. However, future studies with larger samples are necessary to validate the current findings. Moreover, prospective studies are needed in which a treatment selection model is integrated during the diagnostics process at the beginning of a psychotherapy in order to evaluate its added value.

Disclosure statement

The authors report no conflict of interest.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (revised 4th ed.)*. Washington, DC: Author.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421. doi:10.1097/01.psy.0000127692.23278.a9

Backenstrass, M., Pfeiffer, N., Schwarz, T., Catanzaro, S. J., & Mearns, J. (2008). Reliabilität und Validität der deutschsprachigen Version der Generalized Expectancies for Negative Mood Regulation (NMR) Scale [Reliability and validity of the German version of the generalized expectancies for Negative Mood Regulation (NMR) Scale]. *Diagnostica, 54*(1), 43-51. doi:10.1026/0012-1924.54.1.43 .

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., . . . Cuijpers, P. (2016). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *Focus, 14*(2), 229-243.

Berking, M., & Znoj, H. (2008). Entwicklung und Validierung eines Fragebogens zur standardisierten Selbsteinschätzung emotionaler Kompetenzen (SEK-27) [Development and validation of a self-report measure for the assessment of Emotion Regulation Skills (SEK-27)]. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 56*(2), 141-153. doi:10.23668/psycharchives.387

Blom, M. B., Spinhoven, P., Hoffman, T., Jonker, K., Hoencamp, E., Haffmans, P. J., & van Dyck, R. (2007). Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression. *Journal of Affective Disorders, 104*(1-3), 119-126. doi:10.1016/j.jad.2007.03.010

Brähler, E., Mühlan, H., Albani, C., & Schmidt, S. (2007). Teststatistische Prüfung und Normierung der deutschen Versionen des EUROHIS-QOL Lebensqualität-Index und

des WHO-5 Wohlbefindens-Index [Testing and standardization of the German version of the EUROHIS-QOL and WHO-5 quality-of life-indices]. *Diagnostica, 53*(2), 83-96. doi:10.1026/0012-1924.53.2.83

Castonguay, L. G., Eubanks, C. F., Goldfried, M. R., Muran, J. C., & Lutz, W. (2015). Research on psychotherapy integration: Building on the past, looking to the future. *Psychotherapy Research*, 25(3), 365–382. doi:10.1080/10503307.2015.1014010

Castonguay, L. G., Newman, M. G., Borkovec, T. D., Holtforth, M. G., & Maramba, G. G. (2005). Cognitive-behavioral assimilative integration. In J. C. Norcross & M. R. Goldfried (Eds.), *Oxford series in clinical psychology. Handbook of psychotherapy integration* (pp. 241-260). New York, NY, US: Oxford University Press.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., . . . Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry, 3*(3), 243-250. doi:10.1016/S2215-0366(15)00471-X

Clyde, M. A. (2018). BAS: Bayesian Variable Selection and Model Averaging using Bayesian Adaptive Sampling. R package version 1.5.3. Retrieved from https://cran.r-project.org/web/packages/BAS/BAS.pdf

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics, 20*(1), 80-101. doi:10.1198/jcgs.2010.09049

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14*, 209-236. doi:10.1146/annurev-clinpsy-050817-084746

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2019). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to

moderate adult depression. *Psychotherapy Research*, 1-14. doi:10.1080/10503307.2018.1563312

Cuijpers, P. (2014). Personalized treatment for functional outcome in depression. *Medicographia, 36*(4), 476-481.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry, 58*(7), 376-385. doi:10.1177/070674371305800702

Cuijpers, P., Van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological medicine, 40*(2), 211-223. doi:10.1017/S0033291709006114

Cuijpers, P., Van Straten, A., Andersson, G., & Van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of consulting and clinical psychology, 76*(6), 909-922. doi: 10.1037/a0013075

Cuijpers, P. and Dekker, J. 2005. Psychologische behandeling van depressie: een systematisch overzicht van meta-analyses [Psychological treatment of depression: a systematic review of meta-analyses]. *Nederlands Tijdschrift voor Geneeskunde, 149,* 1892–1897.

Deighton, R. M., & Traue, H. C. (2006). Emotionale Ambivalenz, Körperbeschwerden, Depressivität und soziale Interaktion: Untersuchungen zur deutschen Version des “Ambivalence over Emotional Expressiveness Questionnaire “(AEQ-G18) [Emotional ambivalence, illness behavior, depression, and social interaction: Investigations for the German version of the “Ambivalence over Emotional Expressiveness Questionnaire” (AEQ-G18)]. *Zeitschrift für Gesundheitspsychologie, 14*(4), 158-170. doi:10.1026/0943-8149.14.4.158

Deisenhofer, A. K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety, 35*(6), 541-550. doi:10.1002/da.22755

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PloS One, 9*(1), e83875. doi:10.1371/journal.pone.0083875

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., . . . Brown, L. L. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62*(4), 409-416. doi:10.1001/archpsyc.62.4.409

Dinga, R., Marquand, A. F., Veltman, D. J., Beekman, A. T., Schoevers, R. A., van Hemert, A. M., . . . Schmaal, L. (2018). Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Translational Psychiatry, 8*(1), 241. doi:10.1038/s41398-018-0289-1

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician, 37*(1), 36-48. doi:10.2307/2685844

Fernandez, C., Ley, E., & Steel, M. F. (2001a). Benchmark priors for Bayesian model averaging. *Journal of Econometrics, 100*(2), 381-427.

Fernandez, C., Ley, E., & Steel, M. F. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics, 16*(5), 563-576.

Forte, R. M. (2015). *Mastering predictive analytics with R*: Packt Publishing Ltd.

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of

moderate to severe depression. *Journal of Consulting and Clinical Psychology, 77*(4), 775. doi:10.1037/a0015401

Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review, 86*(1), 1-28. doi:10.1111/insr.12243

Greenberg, L. S. (2002). Integrating an emotion-focused approach to treatment into psychotherapy integration. *Journal of Psychotherapy Integration, 12*(2), 154-189. doi:10.1037/1053-0479.12.2.154

Greenberg, L. S. (2017). Emotion-focused therapy of depression. *Person-Centered & Experiential Psychotherapies, 16*(2), 106-117. doi: 10.1080/14779757.2017.1330702

Grosse Holtforth, M. (2008). Avoidance motivation in psychological problems and psychotherapy. *Psychotherapy Research, 18*(2), 147-159. doi:10.1080/10503300701765849

grosse Holtforth, M., & Grawe, K. (2000). Fragebogen zur Analyse motivationaler Schemata (FAMOS) [Questionnaire for the Analysis of Motivational Schemas]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 29*(3), 170-179. doi:10.1026//0084-5345.29.3.170

grosse Holtforth, M., & Grawe, K. (2003). Der Inkongruenzfragebogen (INK) [The incongruence questionnaire (INK). An instrument for the analysis of motivational incongruence]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 32*(4), 315-323. doi:10.1026/0084-5345.32.4.315

grosse Holtforth, M., Hayes, A. M., Sutter, M., Wilm, K., Schmied, E., Laurenceau, J.-P., & Caspar, F. (2012). Fostering cognitive-emotional processing in the treatment of depression: A preliminary investigation in exposure-based cognitive therapy. *Psychotherapy and Psychosomatics, 81*(4), 259-260.

Grosse Holtforth, M., Krieger, T., Zimmermann, J., Altenstein-Yamanaka, D., Dörig, N., Meisch, L., & Hayes, A. M. (2019). A randomized-controlled trial of cognitive–behavioral therapy for depression with integrated techniques from emotion-focused and exposure therapies. *Psychotherapy Research, 29*(1), 30-44. doi:10.1080/10503307.2017.1397796

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine, 363*(4), 301-304. doi:10.1056/NEJMp1006304

Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy of depression: pretreatment patient predictors of outcome. *Clinical Psychology Review, 22*(6), 875-893. doi:10.1016/s0272-7358(02)00106-x

Hautzinger, M., Joormann, J., & Keller, F. (2005). *DAS: Skala dysfunktionaler Einstellungen; Manual [Scale of disfunctional attitudes]*. Göttingen: Hogrefe.

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Das Beck Depressionsinventar II. Deutsche Bearbeitung und Handbuch zum BDI-II [Beck Depressive Inventory II. German Revision and Manual for the BDI-II]*. Frankfurt a.M.: Harcourt Test Services.

Hayes, A. M. (2015). Facilitating emotional processing in depression: The application of exposure principles. *Current Opinion in Psychology, 4*, 61-66. doi:10.1016/j.copsyc.2015.03.032

Hayes, A. M., Beevers, C. G., Feldman, G. C., Laurenceau, J.-P., & Perlman, C. (2005). Avoidance and processing as predictors of symptom change and positive growth in an integrative therapy for depression. *International Journal of Behavioral Medicine, 12*(2), 111-122. doi:10.1207/s15327558ijbm1202_9

Hayes, A. M., Feldman, G. C., Beevers, C. G., Laurenceau, J. P., Cardaciotto, L., & Lewis-Smith, J. (2007). Discontinuities and cognitive changes in an exposure-based cognitive

therapy for depression. *Journal of Consulting and Clinical Psychology*, *75*(3), 409-421. doi: 10.1037/0022-006X.75.3.409.

Hiroe, T., Kojima, M., Yamamoto, I., Nojima, S., Kinoshita, Y., Hashimoto, N., . . . Furukawa, T. A. (2005). Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. *Psychiatry Research, 135*(3), 229-235. doi:10.1016/j.psychres.2004.03.014

Hollon, S. D., Cohen, Z. D., Singla, D. R., & Andrews, P. (2019). Recent Developments in the Treatment of Depression. *Behavior Therapy*. *50*(2), 257-269. doi:10.1016/j.beth.2019.01.002

Huibers, M. J., Cohen, Z. D., Lemmens, L. H., Arntz, A., Peeters, F. P., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PloS One*, *10*(11), e0140771. doi:10.1371/journal.pone.0140771

Hupfeld, J., & Ruffieux, N. (2011). Validierung einer deutschen version der Self-Compassion Scale (SCS-D) [Validation of a German version of the Self-Compassion Scale (SCS-D)]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *40*(2), 115-123. doi:10.1026/1616-3443/a000088

Jerusalem, M., & Schwarzer, R. (1999). Skala zur allgemeinen Selbstwirksamkeitserwartung. *Skalen zur Erfassung von Lehrer-und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen.* Berlin: Freie Universität Berlin.

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review, 67*, 78-93. doi:10.1016/j.cpr.2018.08.004

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Klaghofer, R., & Brähler, E. (2001). Konstruktion und Teststatistische Prüfung einer Kurzform der SCL-90–R [Construction and test statistical evaluation of a short version of the SCL-90–R]. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie, 49*(2), 115-124.

Kraemer, H. C., & Blasey, C. M. (2004). Centring in regression analyses: a strategy to prevent errors in statistical inference. *International Journal of Methods in Psychiatric Research, 13*(3), 141-151. doi: 10.1002/mpr.170

Kuhn, M. (2018). caret: Classification and Regression Training. R package version 6.0-81. Retrieved from https://cran.r-project.org/web/packages/caret/caret.pdf

Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II) [Reliability and validity of the revised version of the Beck Depressive Inventory II]. *Der Nervenarzt, 78*(6), 651-656.

Kühner, C., Huffziger, S., & Nolen-Hoeksema, S. (2007). *Response Styles Questionnaire-Deutsche Version: RSQ-D; Manual*. Göttingen: Hogrefe.

Lampropoulos, G. K. (2001). Bridging technical eclecticism and theoretical integration: Assimilative integration. *Journal of Psychotherapy Integration*, 11(1), 5-19. doi:10.1023/A:102667280

Lemmens, L. H., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huibers, M. J. (2018). Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: follow-up of a randomized controlled trial. *Psychological Medicine, 49*(3), 465-473. doi:10.1017/S0033291718001083

Ley, E., & Steel, M. F. J. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics, 24*(4), 651-674. doi:doi:10.1002/jae.1057

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative Studies of Psychotherapies: Is It True That "Everyone Has Won and All Must Have Prizes"? *JAMA Psychiatry, 32*(8), 995-1008. doi:10.1001/archpsyc.1975.01760260059004

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample Size Requirements for Multivariate Models to Predict Between-Patient Differences in Best Treatments of Major Depressive Disorder. *Clinical Psychological Science*, *7*(3), 445-461. doi:10.1177/2167702618815466

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine, 3*(11), e442. doi:10.1371/journal.pmed.0030442

Messer, S. B. (1992). A critical examination of belief structures in integrative and eclectic psychotherapy. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of Psychotherapy Integration* (pp. 130–165). New York: Basic Books.

Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology, 48*(12), 1503-1510. Doi:10.1016/0895-4356(95)00048-8

Pinheiro, P., Mendes, I., Silva, S., Gonçalves, M. M., & Salgado, J. (2018). Emotional processing and therapeutic change in depression: A case study. *Psychotherapy*, *55*(3), 263–274. doi:10.1037/pst0000190

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111-163. doi:10.2307/271063

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association, 92*(437), 179-191. doi:10.1080/01621459.1997.10473615

Roth, M., Decker, O., Herzberg, P. Y., & Brähler, E. (2008). Dimensionality and norms of the Rosenberg Self-Esteem Scale in a German general population sample. *European Journal of Psychological Assessment, 24*(3), 190-197. doi:10.1027/1015-5759.24.3.190

Röthlin, P., Holtforth, M. G., Bergomi, C., Berking, M., Ottenbreit, N. D., & Caspar, F. (2010). Vermeidung und Depression [Avoidance and Depression]. *Diagnostica*, *56*(1), 46-55. doi:10.1026/0012-1924/a000008

Schulte, D. (2005). Messung der Therapieerwartung und Therapieevaluation von Patienten (PATHEV) [Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV)]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *34*(3), 176-187. doi:10.1026/1616-3443.34.3.176

Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: can we match patients with treatments?. *The American Journal of Psychiatry*, *167*(12), 1445–1455. doi:10.1176/appi.ajp.2010.09111680

Steel, M. F. (2011). Bayesian model averaging and forecasting. *Bulletin of EU and US Inflation and Macroeconomic Analysis*, *200*, 30-41.

Stekhoven, D. J. (2013). missForest: Nonparametric Missing Value Imputation using Random Forest. R package version, 1, 4.

Stekhoven, D. J., & Buhlmann, P. (2012). MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112-118. doi:10.1093/bioinformatics/btr597

Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice, 5*(4), 439-458. doi:10.1111/j.1468-2850.1998.tb00166.x

Stiles, W. B., Shapiro, D. A., & Elliott, R. (1986). Are all psychotherapies equivalent? *American Psychologist, 41*(2), 165-180. doi: 10.1037/0003-066X.41.2.165

Stock, E. M., Kimbrel, N. A., Meyer, E. C., Copeland, L. A., Monte, R., Zeber, J. E., . . . Morissette, S. B. (2014). A Bayesian model averaging approach to examining changes in quality of life among returning Iraq and Afghanistan Veterans. *International Journal of Methods in Psychiatric Research, 23*(3), 345-358. doi:10.1002/mpr.1442

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 10*(6), 363-377. doi:10.1002/sam.11348.

Thase, M. E., Reynolds, C. F., Frank, E., & Simons, A. D. (1994). Response to cognitive-behavioral therapy in chronic depression. *Journal of Psychotherapy Practice & Research, 3*(3), 304-314.

Thomas, A., Brähler, E., & Strauß, B. (2011). IIP-32: Entwicklung, Validierung und Normierung einer Kurzform des Inventars zur Erfassung interpersonaler Probleme [IIP-32: Development, validation, and standardization of a short form of the Inventory of Interpersonal Problems]. *Diagnostica, 57*(2), 68-83. doi:10.1026/0012-1924/a000034

Trösken, A., & Grawe, K. (2002). *Das Berner Ressourceninventar: Konstruktion und Validierung zweier Instrumente zur Erfassung von Patientenressourcen für die klinische Forschung und Praxis. [The Bernese Resource Inventory: construction and validation of two instruments for measuring patient resources for clinical research*

and practice]. Unpublished doctoral thesis. Department of Psychology. University of Bern. Bern, Switzerland.

van Bronswijk, S. C., Lemmens, L. H., Keefe, J. R., Huibers, M. J., DeRubeis, R. J., & Peeters, F. P. (2018). A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder. *Depression and Anxiety*. *36*(3), 252-261. doi:10.1002/da.22868

Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology, 165*(6), 710-718.

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., . . . Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open, 3*(8), e002847.

Wampold, B. E., Minami, T., Baskin, T. W. and Tierney, S. C. 2002. A meta-(re)analysis of the effects of cognitive therapy versus “other therapies” for depression. *Journal of Affective Disorders*, 68: 159–165.

Wampold, B. E., & Brown, G. S. J. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, *73*(5), 914.

Westermann, S., grosse Holtforth, M., & Michalak, J. (2019, in press). Motivation in psychotherapy. In R. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (2nd ed.). New York: Oxford University Press.

World Health Organization. (1995). The World Health Organization Quality of Life Assessment (WHO-QOL): Position paper from the World Health Organization. *Social Science & Medicine, 41*(10), 1403-1409. doi:10.1016/0277-9536(95)00112-K

World Health Organization. (2008). The global burden of disease: 2004 update. Retrieved from https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en

Appendices

Appendix 1

BMA results based on the best 30000 models in the CBT-EE condition

Baseline variables	Prob (%)	Posterior mean (SD)
Sociodemographic variables		
age	99.9	0.2818 (0.07252)
separated/divorced	97.6	-7.627 (0.009889)
apprenticeship	80.6	-2.546 (1.857)
gender	14.1	-0.06219 (0.6017)
single	10.6	-0.09008 (0.009889)
in a relationship/married	6.5	-0.1418 (0.009889)
high school diploma	6.5	0.2333 (1.770)
university diploma	4.6	0.002115 (0.4791)
Depression related variables		
pre-treatment BDI-II	100	-0.7006 (0.1160)
prior antidepressant trials	43.2	1.162 (1.785)
depression severity	17.3	-0.07639 (0.5622)
recurrent depression	7.5	0.02538 (0.5503)
chronic depression	5.5	-0.01286 (0.3953)
depressive episodes	4.8	0.009051 (0.1914)
Other baseline measures		
comorbid axis 2	97.9	-5.180 (2.068)
RRS (self-focused rumination)	97.3	-0.5317 (0.2229)
SCL-K-9	96.6	3.822 (1.674)
comorbid anxiety	93.7	3.527 (2.217)
PATHEV (subscale hopfulness)	91.1	-1.936 (1.080)
DAS subscale performance evaluation	42.5	0.04713 (0.06966)
INC (avoidance)	23.6	0.4941 (1.206)
PATHEV (subscale fear)	23.4	0.2480 (0.6715)
IAAM (avoidance)	21.5	0.5539 (1.460)
SCS (mindfulness)	17.6	-0.2564 (0.8262)
comorbid axis 1	17.4	-0.4674 (1.614)
NMR	17	-0.02505 (0.07659)
IIP dominance	14.2	-0.1319 (0.5501)
INC (approach)	11.9	0.1428 (0.6888)
SCS (over-identification)	11.9	0.1303 (0.6431)
WHO-5	11.5	0.1247 (0.5048)
WHO-QOL	10.6	-0.2408 (1.082)
IIP total score	10.3	0.1864 (0.8249)
RRS (symptom focused rumination)	10	-0.01843 (0.09745)
IAAM (approach)	9.5	0.00006262 (0.0007448)
CBAS	9.4	0.1117 (0.6219)
SCS (isolation)	9	0.04062 (0.4122)
SWE	9	-0.05152 (0.5938)
DAS subscale recognition by others	8.7	-0.001500 (0.03221)
IIP love	8.6	0.04659 (0.4.117)
SCS (self-kindness)	8.6	0.08874 (0.4961)
SEK-27	7.4	-0.00382 (0.02165)
RSES	6.9	0.01247 (0.4049)
AVEX	5.9	-0.002390 (0.04782)
SCS (self-judgment)	5	0.02540 (0.3322)
SCS (common humanity)	3.8	0.01751 (0.2800)
RES	3.5	0.01916 (0.2904)

Appendix 2

BMA results based on the best 30000 models in the CBT condition

Baseline variables	Prob (%)	Posterior mean (SD)
Sociodemographic variables		
gender	50.8	2.030 (2.524)
age	10.5	-0.003545 (0.03373)
university diploma	9.3	0.002596 (0.7654)
in a relationship/married	8.7	0.06092 (0.1095)
apprenticeship	7.3	0.1312 (0.8016)
separated/divorced	6.6	-0.1349 (0.1095)
single	5.7	0.02171 (0.1095)
high school diploma	5.3	-0.06843 (1.525)
Depression related variables		
pre-treatment BDI-II	100	-0.7702 (0.1444)
recurrent depression	83.3	7.868 (5.390)
depressive episodes	81.1	-2.461 (1.816)
chronic depression	20.1	-0.2829 (1.220)
prior antidepressant trials	10.1	0.09598 (1.032)
depression severity	7.8	-0.05168 (0.5032)
Other baseline measures		
INC (avoidance)	67.6	2.205 (2.128)
WHO-QOL	43.9	-1.266 (2.364)
comorbid anxiety	20.2	-0.7581 (2.030)
CBAS	16.9	-0.1956 (0.9814)
comorbid axis 2	14.8	0.2960 (1.119)
RSES	14.1	-0.3169 (1.033)
SCS (self-kindness)	13.3	-0.2281 (0.8699)
PATHEV (subscale fear)	11.7	0.1301 (0.6904)
RES	11	0.08776 (0.5777)
NMR	10.2	0.007748 (0.05352)
SCS (common humanity)	9.9	0.1382 (0.7022)
DAS subscale performance evaluation	9.5	-0.006471 (0.03210)
IIP total score	8.9	0.02489 (0.8316)
SCS (isolation)	8.9	-0.1445 (0.7056)
AVEX	8.8	-0.01103 (0.07968)
SCS (over-identification)	8.7	0.06812 (0.5919)
IIP love	8.6	0.1214 (0.6034)
WHO-5	8.4	-0.03642 (0.5178)
IIP dominance	7.8	-0.02376 (0.4832)
SCL-K-9	7.5	0.09329 (0.7251)
SCS (self-judgment)	7.4	0.05268 (0.5729)
SWE	7.3	-0.1011 (0.8388)
comorbid axis 1	7.1	0.03592 (1.258)
RRS (symptom focused rumination)	6.5	0.003873 (0.09040)
RRS (self-focused rumination)	6.5	-0.001968 (0.08079)
SCS (mindfulness)	6.2	-0.02664 (0.4927)
INC (approach)	5.9	0.03923 (0.7097)
PATHEV (subscale hopfulness)	5.6	-0.03799 (0.5134)
DAS subscale recognition by others	5.6	0.001.493 (0.03634)
SEK-27	5.4	0.001799 (0.01931)
IAAM (avoidance)	5.2	-0.04496 (0.6814)
IAAM (approach)	4.7	-0.02718 (0.6353)

2.2 Study 2: Using the Personalized Advantage Index for individual treatment allocation to blended treatment or treatment as usual for depression in secondary care

2.2.1 Goal of the study and research questions

Similar to Study 1, Study 2 uses the PAI approach for treatment selection for patients with depression. Blended treatment was compared to treatment as usual in secondary care in four European countries. Blended treatment is a new approach to treatment for depression that combines face-to-face therapies with internet-based elements. Research suggests the feasibility and efficacy of blended treatment and even though predictive modeling in e-mental health is still very young, machine learning techniques have been used in prior studies (Bremer et al., 2018). Study 2 had the goal to identify the important predictors of treatment outcome in both conditions and to investigate whether model-determined treatment allocation using predictive information results in better treatment outcomes.

2.2.2 Sample of Study 2

The data for Study 2 was drawn from the European project “European COMPARative Effectiveness research on blended Depression treatment” (E-COMPARED) (Kleiboer et al., 2016). This project investigated the clinical and cost-effectiveness of blended treatment compared to treatment as usual in routine care in nine European countries using a randomized controlled, non-inferiority trial. The current study used data from the Netherlands (N=83, 33.9%), France (N=79, 32.2%), Switzerland (N=44, 18%) and Denmark (N=39, 15.9%) and has a total sample size of N=245. 68.2% of the participants were female and the mean age at baseline was 41.0 years (SD= 13.65). 33.5% were single, 31.8% were married and 12.8% were divorced. In the TAU condition, 57.9% of the sample suffered from a recurrent depression, 46.8% from a comorbid anxiety disorder and 53.2% of the patients are taking

antidepressant medication at the time of the baseline measurement. In the blended treatment condition, 53.8% of the participants suffered from a recurrent depression, 61.3% had a comorbid anxiety disorder and half of the participants (49.6%) were taking antidepressant medication at baseline.

Patients who were randomized to blended treatment received individual face-to-face CBT with CBT-elements delivered through an internet-based treatment platform. The number of face-to-face sessions was reduced and replaced by online modules. The TAU treatment was defined as the treatment that patients with a diagnosis of depression receive in specialized mental health care, thus regular face-to-face CBT.

2.2.3. Data analytical strategy of Study 2

The primary outcome measure was the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001) assessed after 12 weeks. Cronbach's alpha was .78. A total of 28 potential predictors that were measured at baseline have been included and can be classified into (1) sociodemographic variables, (2) symptomatology and quality of life, (3) healthcare utilization and (4) patient expectancy. Quality of life was measured with the EQ-5D (EuroQol, 1990) and healthcare utilization was assessed with the TiC-P which is a self-report questionnaire that examines healthcare consumption (Bouwmans et al., 2013). The Credibility and Expectancy Questionnaire (CEQ) was used to measure patient expectancy (Devilly & Borkovec, 2000).

After removing missing baseline data with the R package *missForest* (Stekhoven, 2013), the data set was split into two subsets, the face-to-face condition and the blended treatment condition. Then, BMA (Fragoso et al., 2018) was used for variable selection with treatment outcome as the dependent variable and to compute separate linear regression models for each treatment condition. Posterior probabilities were used to evaluate the relative

importance of each potential predictor. Variables with a posterior probability over 0.5 are defined as important because they are included in over 50% of all the models. In total, 30,000 linear regression models were estimated. For generating the PAI, regression models using a leave-one-out approach (Efron & Gong, 1983) were estimated for each patient. This means that each patient for whom the PAI prediction is estimated is excluded from the model to avoid overfitting. For each patient, a *factual prediction* (post-treatment PHQ-9 score of the treatment the patient has received) and a *counterfactual prediction* (post-treatment PHQ-9 score of the intervention the patient did not receive) were estimated. Finally, the PAI results from the predicted difference of receiving the treatment with the greater predicted benefit.

2.2.4. Results of Study 2

Variables predicting outcome in TAU

The most important variables predicting the PHQ-9 score at 12 weeks in the TAU condition were the pre-treatment PHQ-9 score (Prob=100%), CEQ expectancy (Prob=97%), country of treatment Denmark (Prob=58%) and the following items from the TiC-P: “How many days did you use outpatient psychotherapeutic services in addition to psychotherapy?” (Prob=95%), “How many times did you consult a psychiatrist?” (Prob=64%) and “How many days did you attend a day-time treatment program in a psychiatric hospital?” (Prob=51%). A higher pre-treatment score, more consultations with a psychiatrist and more days in a day-time treatment program in a psychiatric hospital prior to treatment predicted a higher PHQ-9 score at 12 weeks. Higher expectancy scores, receiving TAU in Denmark, and more days using outpatient psychotherapeutic services in addition to the psychotherapy prior to treatment predicted lower PHQ-9 scores at 12 weeks.

Variables predicting outcome in blended treatment

The most important predictors of PHQ-9 scores at 12 weeks in the blended treatment condition were pre-treatment PHQ-9 scores (Prob=99.9%), regular hospital admissions (Prob=99.9%), EQ-5D quality of life (Prob=74.6%), CEQ expectancy (Prob=72.3%), consulting self-help groups (Prob=70%) and being widowed (Prob=49.7%). A higher pre-treatment PHQ-9 score, being widowed, more hospital admissions and consulting self-help groups predicted higher PHQ-9 scores at 12 weeks. A higher expectancy for improvement and a higher quality of life predicted lower PHQ-9 scores after 12 weeks.

The PAI

The true error of the PHQ-9 predictions at 12 weeks was 4.16 which represents the difference between the predicted and actual scores across all patients. Patients who were categorized as having received their optimal treatment had a mean PHQ-9 score of 9.67 (n = 124) at 12 weeks, whereas patients who were classified as having received their suboptimal treatment had a mean PHQ-9 score of 12.00 (n = 121). This results in an average PAI of 2.33 points on the PHQ-9. Consequently, if patients had received their model-determined optimal treatment, their PHQ-9 score at 12 weeks would have been 2 points lower than if they had received the suboptimal treatment.

2.2.5 Manuscript of Study 2

Friedl, N., Krieger, T., Chevreul, K., Hazo, JB., Holtzmann, J., Hoogendoorn, M., Kleiboer, A., Mathiasen, K., Urech, A., Riper, H., & Berger, T., (in press). Using the Personalized Advantage Index for individual treatment allocation to blended treatment or treatment as usual for depression in secondary care.

Using the Personalized Advantage Index for individual treatment allocation to blended treatment or treatment as usual for depression in secondary care

Nadine Friedl [1,*], Tobias Krieger [1], Karine Chevreul [2], Jean Baptiste Hazo [3], Jérôme Holtzmann [4], Mark Hoogendoorn [5], Annet Kleiboer [6], Kim Mathiasen [7,8], Antoine Urech [9], Heleen Riper [10,11] and Thomas Berger [1]

[1] Department of Clinical Psychology and Psychotherapy, University of Bern, Switzerland
[2] URC-ECO, Ile-de-France (AP-HP), Paris, France
[3] Unité de Recherche en Economie de la Santé, Assistance Publique - Hôpitaux de Paris, Paris, France
[4] Unité troubles de l'humeur et pathologies des émotions, Pôle de psychiatrie, neurologie et rééducation neurologique, Centre Hospitalier Universitair Grenoble Alpes, France
[5] Faculty of Sciences, Department of Computer Science, Vrije Universiteit Amsterdam, The Netherlands
[6] Section Clinical Psychology, Vrije Universiteit Amsterdam and EMGO+ Institute for Health Care and Research, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands
[7] Department of Psychology, Faculty of Health Sciences, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark
[8] Centre for Telepsychiatry, Mental Health Services of Southern Denmark, Odense, Denmark
[9] INSELSPITAL, Universitätsspital Bern, Universitätsklinik für Neurologie, Universitäre Akut-Neurorehabilitation; Bern, Switzerland
[10] Department of Research and Innovation, GGZ inGeest, Specialized Mental Health Care, 1081 Amsterdam, Netherlands
[11] Department of Psychiatry, Amsterdam Public Health Research Institute, 1081 Amsterdam UMC, Vrije Universiteit, Amsterdam

*Adress for correspondence
Department of Clinical Psychology and Psychotherapy, University of Bern
Fabrikstrasse 8, CH-3012 Bern, Switzerland
Tel: +41 (0) 31 631 36 91
E-Mail: nadine.friedl@psy.unibe.ch

Abstract: A variety of effective psychotherapies for depression are available but patients who suffer from depression vary in their treatment response. Combining face-to-face therapies with internet-based elements in the sense of blended treatment is a new approach to treatment for depression. The goal of this study was to answer the following research questions: (1) What are the most important predictors determining optimal treatment allocation to treatment as usual or blended treatment?, and (2) Would model-determined treatment allocation using this predictive information and the Personalized Advantage Index (PAI)-approach result in better treatment outcomes? Bayesian Model Averaging (BMA) was applied to the data of a randomized controlled trial comparing the efficacy of treatment as usual and blended treatment in depressive outpatients. Pre-treatment symptomatology and treatment expectancy predicted outcome irrespective of treatment condition, whereas different prescriptive predictors were found. A PAI of 2.33 PHQ-9 points was found, meaning that patients who would have received the treatment that is optimal for them would have had a post treatment PHQ-9 score that is two points lower than if they had received the treatment that is suboptimal for them. For 29% of the sample, the PAI was five or greater which means that a substantial difference between the two treatments was predicted. The use of the PAI approach for clinical practice must be further confirmed in prospective research, the current study supports the identification of specific interventions favorable for specific patients.

Keywords: Personalized Advantage Index; depression; blended treatment; CBT; treatment selection; Bayesian Model Averaging

1. Introduction

Globally, 300 million people of all ages suffer from depression [1]. Depression is one of the most common problems seen in clinical practice and it is associated with high societal costs as well as great suffering [2]. Given this burden, the need for improved access to efficacious and cost-effective treatments is essential [3]. In the last decades, research has focused on examining different treatment options for depression. Especially cognitive behavior therapy (CBT) and interpersonal therapy (IPT) can be seen as first-line treatments [4, 5]. Moreover, current studies are aimed at scaling up treatment for depression. One way to do so is through internet-based therapies [5]. Whereas the most dominant format in which treatment is delivered is through face-to-face contact, internet-based therapies have received much attention in recent years [6]. The efficacy and cost-effectiveness of the latter have been supported by a growing number of research [7-9]. Even though only a few studies have directly compared internet-based with face-to-face CBT for depression, results suggest it to have similar overall effects [6].

A newer approach to depression treatment is to combine web-based technologies with face-to-face therapy, called *blended treatment*. *Blended treatment* includes any combination of face-to-face therapy and internet-based interventions, e.g., web-based components are used as an adjunctive intervention or are integrated during face-to-face therapy [10]. Although research that investigates the efficacy of blended treatment formats is still scarce, preliminary results suggest their feasibility and their efficacy in reducing symptoms of depression [11-15]. For example, a randomized controlled trial by Berger and colleagues [16] showed the superiority of blended treatment, consisting of an internet-based intervention as an adjunct to face-to-face psychotherapy, in comparison to regular face-to-face psychotherapy in a pragmatic randomized controlled study in patients with a unipolar affective disorder in routine care. Another recent study showed the noninferiority of blended treatment to conventional CBT for patients with

depression and found blended treatment to be cost-effective [17]. Moreover, blended treatment has also been evaluated in an inpatient setting where patients suffering from depression that received an online self-help program in addition to inpatient psychotherapy improved significantly more than patients who received online information about depression in addition to inpatient psychotherapy [18]. Furthermore, a recent systematic review showed that compared to face-to-face therapy, blended treatment may help maintaining initially achieved changes within psychotherapy in the long-term [19].

Potential benefits of blending treatments may be a greater reduction in depressive symptoms and increased cost-effectiveness [20] as well as an improvement of patients' adherence to the treatment program [21]. Furthermore, an asset of blended treatment may be that it combines the advantages of both treatment forms [3, 14, 22, 23]. For example, the face-to-face contact enables clinicians to individualize or tailor the treatment and to react in crisis situations, while providing online modules between sessions could promote patient engagement and enhance the translation of treatment into daily life (e.g., [24]). On the other hand, when online components are not used by the patients in blended treatments, reductions in the number of face-to-face sessions may lead to worse treatment outcomes [25]. Furthermore, therapists may raise concerns of overburdening depressed patients [18]. So far it is not clear, for which patients blended treatment may be a feasible option and for which patients a conventional treatment should be favored.

Patients with depression may differ substantially from each other and evidence suggests that the diagnostic categories leave room for great diversity [26, 27]. This results in differences with regard to patients' illness course and individual treatment response [28]. Research suggests that individual patient characteristics may moderate the efficacy of different treatments at an individual level [29]. It is, therefore, important to recognize that no single treatment is likely to be the best for everyone, even though on a group level, it is efficacious for patients suffering

from depression [26, 30]. This is why more and more researchers move away from investigating treatment efficacy on a group level and instead focus on custom-tailoring the treatment to the individual patient [30, 31]. In this sense, it may be a solution to increase the overall treatment response rates [32].

Precision medicine tries to tailor treatments to the specific needs of the patient [33]. More recently, in clinical psychology and psychotherapy, algorithms are used that predict from which treatment a patient benefits the most [34]. As an example, Becker and colleagues [35] introduce a conceptual framework that helps classifying applications of predictive modeling in mental health research. These authors try to bridge the gap between psychologists and predictive modelers with providing a common language for classifying predictive modeling mental health research. They suggest that e-mental health researchers should focus more on the validity of model predictions instead of solely focusing on identifying predictors. Another example is the *Personalized Advantage Index* approach, which identifies patients with a certain disorder (e.g., major depression) who benefit more from one treatment than another [30]. Using the Personalized Advantage Index (PAI), it is possible to identify the treatment that predicts a better treatment outcome for a given patient if there are two comparably effective treatments to choose from [36]. The PAI estimates how much a specific treatment is better for an individual patient than another, and its feasibility and relevance have been shown in several studies on the treatment of depression [36-40]. Baseline patient characteristics can be can be divided into two types of predictors: A *prognostic* variable predicts treatment outcome irrespective of treatment condition whereas a *prescriptive* variable predicts a differential treatment response to two or more treatment modalities [29, 30]. Up to today, different treatments have been compared using the PAI and its values range from 1.4 when comparing CBT to CBT with integrated exposure and emotion-focused elements [38] up to 8.9 when comparing Cognitive Therapy to IPT [40]. Higher absolute values of the PAI stand for stronger predicted benefits of one treatment over

another. Being able to identify the best treatment for an individual with depression is essential because it may make health care delivery more efficient [32]. Even though predictive modeling is still very young in the field of e-mental health [41], Bremer and colleagues [42] were able to predict clinical outcomes and costs of patients with depression prior to starting blended psychotherapy in a subsample of the current study using machine learning techniques.

In the current study, treatment as usual (TAU), i.e., regular face-to-face psychotherapy, was compared to blended treatment for patients with major depressive disorder (MDD) in secondary care. The present study was set out to answer the following research questions: (1) What are the most important predictors determining optimal treatment allocation to TAU or blended treatment?, and (2) Would model-determined treatment allocation using this predictive information and the PAI-approach result in better treatment outcomes? To the best of our knowledge, this is the first study comparing different treatment delivering formats, i.e., traditional face-to-face CBT versus blended CBT, by using the PAI-approach.

2. Materials and Methods

Data used in the present study was drawn from the European project “European COMPARative Effectiveness research on blended Depression treatment” (E-COMPARED, February 2018) [43]. The E-COMPARED project included a randomized controlled, noninferiority trial that examined the clinical and cost-effectiveness of blended treatment compared to treatment as usual in routine care in nine European countries. Adult patients diagnosed with MDD were recruited in primary or in specialized mental health care. The current study uses the data of the four countries that recruited patients in specialized mental health care (France, the Netherlands, Switzerland and Denmark). Following inclusion criteria were met by participants: (1) being 18 years of age or older, (2) meeting Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) diagnostic criteria for MDD as confirmed by a telephone-administered MINI International Neuropsychiatric Interview (M.I.N.I.) version 5.0 [44] and (3)

minimal to severe symptoms of depression based on a score of 5 or above on the Patient Health Questionnaire-9 (PHQ-9) [45]. Exclusion criteria for participating in the study were: (1) high risk for suicide according to the M.I.N.I., (2) psychiatric comorbidity such as substance dependence, bipolar affective disorder, psychotic illness, obsessive compulsive disorder, (3) currently receiving another psychological treatment for depression, (4) being unable to comprehend the spoken and written language of the country where the study is conducted, (5) not having access to a computer with fast Internet connection and (6) not having a smartphone that is compatible with the mobile component of the intervention that is offered. Patients were randomized to blended treatment or TAU using an allocation scheme with a computerized random number generator at an allocation ratio of 1:1 and between 8 and 14 allocations per block. Details about the study design are described elsewhere [43]. For Switzerland, this study was approved by the cantonal ethics committees Bern and Zurich (registration number: 001/15; date 18/03/2015), for Denmark, the study was approved by the Ethics Committee of the Region of Southern Denmark (registration number S-20150150; date 18/11/2015), for France, the study was approved by the "Comité de protection des personnes", Ile de France V (registration number 15033-n° 2015-A00565-44; date 2/06/2015) and for the Netherlands, the study was approved by METC VUMC (registration number 2015.078; date 8/05/2015. Furthermore, all participants provided written informed consent and gave permission to all E-COMPARED partners to use their anonymized data.

2.1. Sample

The current study has a sample size of $N = 251$. The sample consists of 83 participants from the Netherlands (33.9%), 79 participants from France (32.2%), 44 from Switzerland (18.0%), and 39 participants from Denmark (15.9%). The mean age at baseline was 41.0 years (SD = 13.7) and 68.2% of the participants were female. The majority were either single (33.5%) or married (31.8%), and 21.2% of participants were living together and 12.8% were divorced. In

the TAU condition, 57.9% of the sample suffered from a recurrent depression, 45.2% from a current melancholic depressive episode, 7.9% from a comorbid dysthymia and 46.8% from a comorbid anxiety disorder. 53.2% of the patients are taking antidepressant medication at the time of the baseline measurement. In the blended treatment condition, 53.8% of the participants suffered from a recurrent depression, 34.5% from a current melancholic depressive episode, 5.9% from a comorbid dysthymia and 61.3% had a comorbid anxiety disorder. Half of the participants (49.6%) were taking antidepressant medication at baseline.

2.2. Interventions

Individual face-to-face CBT was combined with internet-based CBT-elements delivered through a platform for blended treatment. Three different online platforms were used across the participating countries. Switzerland used “Deprexis” [46], whereas Denmark used NoDep [47] and France and the Netherlands used the platform “Moodbuster” [3]. The most important components of the treatment were cognitive restructuring, behavioral activation, psychoeducation, and relapse prevention that were delivered over 11-20 sessions. In the blended treatment, a smaller number of face-to-face sessions is offered and some sessions are replaced by online modules. Treatment was provided by CBT therapists who received special training on how to deliver blended treatment. In Switzerland and the Netherlands, the therapists were either licensed CBT therapists or CBT therapists who were supervised by an experienced licensed CBT therapist. In Denmark, the treatment was delivered by either licensed psychologists or psychologist under supervision of licensed psychologists. In France, blended treatment was provided by licensed psychotherapists [43].

The TAU treatment was defined as the routine care that patients received in specialized mental health care when they are diagnosed with depression. In practice, this meant that the TAU group received regular fact-to-face CBT.

2.3. Measures

2.3.1. Primary outcome

The primary outcome measure for this study was the PHQ-9 [45] assessed after 12 weeks. The PHQ-9 consists of nine questions which are based upon the DSM-IV criteria for the diagnosis of depressive disorders. It is used as a diagnostic instrument and as a severity measure for depression. A 5-point difference in PHQ-9 scores is seen as clinically significant [48]. The validity and sensitivity to change of the PHQ-9 were satisfactory in previous studies [49] [50]. Cronbach's alpha in the present study was .78.

2.3.2. Predictor variables

We used an exploratory approach and included a total of 28 potential predictors measured at baseline in the analysis. All variables of the baseline assessment that did not exceed a number of acceptable missing values (< 50%) were included. We classified the variables into four categories: (1) sociodemographic variables, (2) symptomatology and quality of life, (3) healthcare utilization and (4) patient expectancy.

The sociodemographic variables included age, marital status, education, gender and country and were assessed with single item questions. Variables related to symptomatology and treatment history were recurrent depression, therapy preference, dysthymia, melancholic depressive episodes, comorbid anxiety, and current use of antidepressants. Quality of life was measured with the EQ-5D [51]. Healthcare utilization was assessed with the TiC-P [52]. The TiC-P examines healthcare consumption and productivity losses as a consequence of a mental disorder via a self-report questionnaire. The questions include contacts within the healthcare sector and the use of medication. All the questions aim at the period of the last four months before the start of the treatment (see Table 1 for the items of the TiC-P that have been included).

Patient expectancy was measured by the Credibility and Expectancy Questionnaire (CEQ; [53]).

Table 1

TiC-P items included in the analysis

"How many times did you consult a general practitioner?", "How many times did you consult a psychologist?", "How many times did you consult a psychotherapist?", "How many times did you consult a psychiatrist?", "How many times did you consult a professional from an ambulatory mental health institution?", "How many times did you consult a professional from a clinic for alcohol or drugs?", "How many times did you consult self-help groups?"	"How many days did you spend in a day-time treatment program in a regular hospital?", "How many days did you spend in a day-time treatment program in a psychiatric hospital?", "How many days did you use outpatient psychotherapeutic services in addition to your psychotherapy?"	"How many admissions to a regular hospital did you have?", "How many admissions to a psychiatric hospital did you have?", "How many admissions to a rehabilitation clinic did you have?", "Do you have a paid job?", "Did health problems oblige you to call in sick from work at any time?"

2.4. Data analytical strategy

Regarding the predictor variables, a bottom-up approach was followed which means that even though some variables might have a particular relevance to one treatment or the other, the predictors are treated equally in the data analysis.

2.4.1. Missing data

In line with previous research using the PAI approach, we included those participants for which PHQ-9 scores after 12 weeks were available [36, 38, 40]. This left us with N = 245, representing 97.6% of the total sample. Distributed over the two conditions, there were 126 patients in the TAU condition and 119 patients in the blended treatment condition. With regard to the baseline measures, missing values were found in the dimensions credibility and expectancy of the CEQ (3.7% resp. 4.1%), the EQ-5D (1.6%), antidepressants (1.2%), prior psychotherapy (40.4%), comorbid melancholic episodes (14.3%), comorbid dysthymia (5.3%), comorbid anxiety (2.4%), and some items of the TiC-P (2.0-40.8%). We imputed missings in the baseline measures with the R *missForest* [54]. Here missing values are predicted on the basis of a random forest approach, trained on observed values of the available data. An advantage of *missForest* is that it imputes categorical and continuous variables simultaneously [55]. It has been shown to be highly accurate and to outperform other imputation methods due to its small imputation error [56]. The imputed data set is the basis for all analyses that follow.

2.4.2. Bayesian Model Averaging (BMA)

There are different data analysis approaches that can be used to identify baseline variables that predict outcome in one treatment versus another. Data analysis approaches that rely on one model only, can only include a limited number of baseline measures that could be predictors. In that case, the most common rule that is used states that at least 10 observations per predictor are necessary to not exceed a level of bias which is acceptable [57, 58]. If we had followed this

approach, we could have only included a small number of predictors. Furthermore, the problem with relying upon a single selected model is that it may result in overconfidence in the conclusions drawn regarding quantified associations. The problem is that there may be alternative models that have different subsets of predictors that fit the data just as well as the one selected model [59, 60]. Overfitted models cause uncertainty regarding the actual value of findings because they may not replicate in future samples [61]. Bayesian Model Averaging (BMA) is a method that can account for model uncertainty while providing a better predictive ability. The BMA method has two advantages. First, it results in predictions that are less risky and second, BMA provides simpler model choice criteria because it uses Bayesian inference for model prediction and selection [62]. With BMA, a posterior probability is estimated on the basis that each considered model is the correct one. This included the aforementioned model uncertainty in the estimates for the parameter and inferences. That means that BMA averages over all possible predictor sets and delivers model choice criteria that help to identify the most probable model. With using BMA, sharper predictions can be derived from the data, especially in cases with many possible predictors but a limited sample size. Several studies have supported BMA's predictive performance [60, 63-66].

The data frame was devided into two subsets, the TAU condition, and the blended treatment condition. Using the R package *BAS*, for each condition, a separate linear regression model was computed [67]. Then Bayesian adaptive sampling (BAS) without replacement for variable selection in linear models using the function *bas.lm* with treatment outcome was applied (PHQ-9 score after 12 weeks) as the dependent variable. The relative importance of each variable is evaluated using the posterior probabilities that are calculated for each potential predictor. The marginal posterior inclusion probabilities function as the criteria for determining the importance of the potential predictors. Values above 0.5 point out that the predictor has been incorporated in more than half the models, thus in the present study in over 15,000 models. The nominal

variables are split into their different categories, which enables a precise interpretation of the results. The model performance was further evaluated using posterior probabilities. The appropriate *bas.lm* function was chosen based on the following considerations: a Laplace approximation to the *Jeffreys-Zellner-Siow* (JZS) prior for the integration of alpha = 1 was used as the criterion for the priors which is called the *"ZS-null"*. The *JZS prior* uses the Zellner-Siow Cauchy prior on the coefficients and the Jeffreys prior on sigma. The squared scale of the prior, where the default is alpha=1 can be controlled using the optional parameter 'alpha' [38, 67]. Marginal inclusion probabilities were calculated with the *"MCMC+BAS" method*, which runs an initial Markov chain Monte Carlo (MCMC) algorithm and then samples without replacement as in BAS. Compared to the BAS alone, the “MCMC+BAS” method is the preferred option because it provides estimates with low bias [68]. The number of models was set to 30,000 assuming that each additional model would add only a small increment to the cumulative probability, i.e., not leading to essential differences in posterior distributions. Due to the limited sample size, the models have been built and tested in the same data set.

2.4.3. Personalized Advantage Index (PAI)

When predicting the therapy outcome for each patient, applying a leave-one-out approach (“jackknife”) to estimate regression models is an essential beginning step in generating the PAI [36, 38, 69]. In this procedure, overfitting can be avoided by excluding each target patient for whom the PAI prediction is estimated from the model. For each patient, two regression models were built using the treatment specific predictors identified with BMA. For each patient, a *factual prediction* (PHQ-9 score at 12 weeks of the treatment the patient has received) and a *counterfactual prediction* (PHQ-9 score at 12 weeks of the intervention the patient did not receive) were estimated. In the next step, those two predictions were compared and the prediction that resulted in the best outcome for the patient was defined to be the optimal treatment for that patient. When comparing the observed change scores, the size of the predicted

difference of receiving the treatment with the greater predicted benefit is ultimately the PAI [36]. The higher the absolute values of the PAI, the stronger is the predicted benefit of one treatment over another. The interpretation of the PAI can be demonstrated with a recent study that used the PHQ-9 as primary outcome and found a PAI of 2.5 [70]. This means that if patients had received their "optimal" treatment (out of the two), their PHQ-9 score at 12 weeks would have been 2.5 points lower than if they had obtained their non-optimal treatment.

3. Results

Henceforth, we firstly report the five best models for each treatment condition, and secondly, we report the PAI results. The best models are defined based on the highest posterior probability and the lowest BIC.

3.1. Variables predicting outcome in TAU

The five best models predicting depression severity at 12 weeks in the TAU condition are displayed in Table 2. The Bayes Factor, number of predictors, R^2, log marginal likelihood, and the posterior probabilities are provided for each model. Model 1 has the largest Bayes Factor and the largest posterior probability (0.02) and thus seems to fit the data best. As a result, Model 1 was selected as our final predictive model of the PHQ-9 score at 12 weeks in the TAU condition.

Table 2

Five best models TAU

Fit indices	Model 1	Model 2	Model 3	Model 4	Model 5
Bayes Factor	1	0.934	0.597	0.572	0.510
Number of variables	6	6	6	5	6
R^2	0.428	0.445	0.441	0.440	0.439
log marginal likelihood	22.729	22.659	22.213	22.170	22.056
Posterior probabilities	0.021	0.019	0.012	0.012	0.011

While the selected model includes six variables in total, the strongest predictors of the PHQ-9 score at 12 weeks in the TAU condition included pre-treatment PHQ-9 score (Prob=100%), CEQ expectancy (Prob=97%), “How many days did you use outpatient psychotherapeutic services in addition to your psychotherapy?”, (Prob=95%), “How many times did you consult a psychiatrist?” (Prob=64%), Denmark (Prob=58%) and “How many days did you attend a day-time treatment program in a psychiatric hospital?” (Prob=51%). A higher pre-treatment score, more consultations with a psychiatrist and more days in a day-time treatment program in a psychiatric hospital prior to treatment predicted a higher PHQ-9 score at 12 weeks. Higher expectancy scores, receiving TAU in Denmark, and more days using outpatient psychotherapeutic services in addition to the psychotherapy prior to treatment predicted lower PHQ-9 scores at 12 weeks. The effects of other variables appeared minimal due to their small posterior probabilities. See Appendix 1 in the supplemental online material for the complete list of variables and their inclusion probabilities.

3.2. Variables predicting outcome in the blended treatment

Table 3 gives an overview of the five best models to predict treatment outcome in the blended treatment condition. Based on the posterior probabilities and the Bayes Factor, Model 1 was rated the best model. Thus, Model 1 was selected as the final predictive model for the blended treatment condition.

Table 3

Five best models blended treatment

Fit indices	Model 1	Model 2	Model 3	Model 4	Model 5
Bayes Factor	1	0.887	0.577	0.525	0.516
Number of variables	6	5	4	6	6
R^2	0.398	0.416	0.392	0.429	0.446
log marginal likelihood	19.852	19.731	19.301	19.207	19.190
Posterior probabilities	0.012	0.011	0.007	0.006	0.006

Based on the posterior probabilities, the most important predictors for treatment outcome in blended treatment were pre-treatment PHQ-9 score (Prob=99.9%), regular hospital admissions (Prob=99.9%), EQ-5D quality of life (Prob=74.6%), CEQ expectancy (Prob=72.3%), consulting self-help groups (Prob=70.0%) and being widowed (Prob= 49.7%). CEQ credibility reached a posterior probability of 42.9%. A higher pre-treatment PHQ-9 score, being widowed, more hospital admissions and consulting self-help groups predicted higher PHQ-9 scores at 12 weeks. A higher expectancy for improvement and a higher quality of life predicted lower PHQ-9 scores after 12 weeks. See Appendix 2 in the supplemental online material for the posterior probabilities of all variables measured at baseline.

3.3. Personalized Advantage Index

Using the treatment specific predictors described above, the prediction of a patient's PHQ-9 score after 12 weeks was computed separately for each treatment condition. The true error of the PHQ-9 score predictions at 12 weeks was 4.16, representing the average absolute difference between the predicted and actual, observed scores across all patients. Patients who were categorized as having received their optimal treatment had a mean PHQ-9 score of 9.67 (n = 124) at 12 weeks, whereas patients who were classified as having received their suboptimal

treatment had a mean PHQ-9 score of 12.00 (n = 121). Figure 1 shows the frequency of predicted PHQ-9 scores at 12 weeks for every patient in both the optimal and suboptimal treatment.

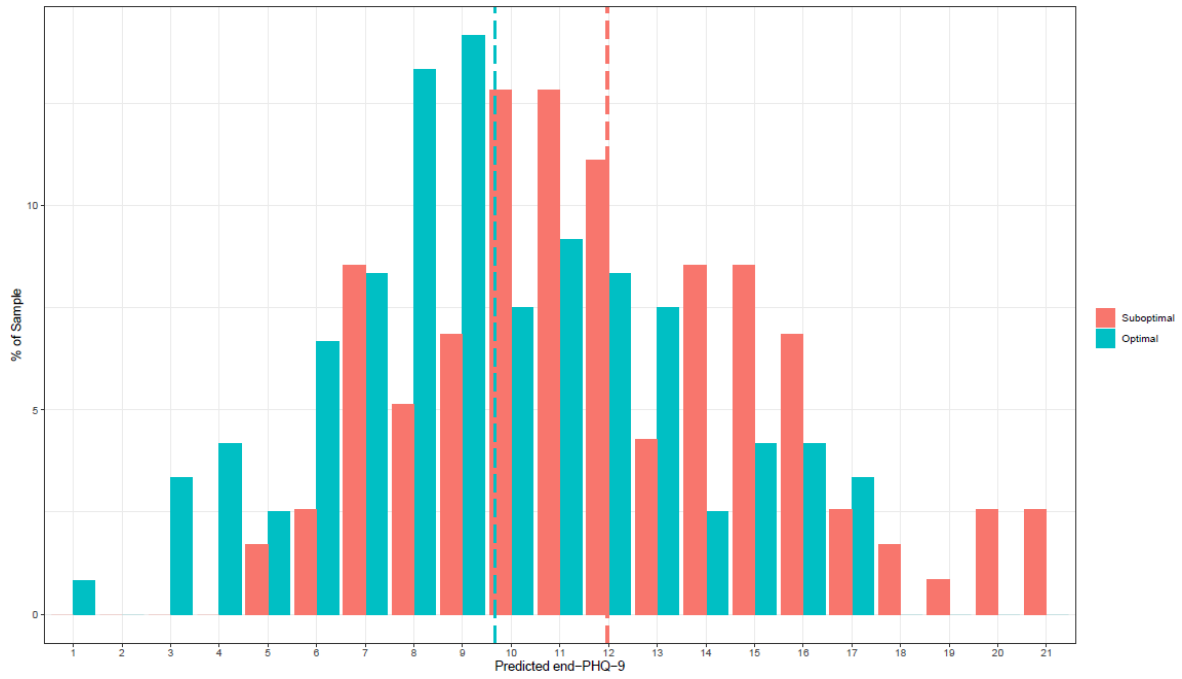

Figure 1. Frequency of predicted PHQ-9 scores at 12 weeks.

In the first step, an individual PAI was calculated for each patient. Secondly, the average PAI was calculated as the mean difference in PHQ-9 scores between the optimal and the suboptimal treatment for each patient. The average PAI of the current study is 2.33. The PAI can be read as follows: if patients had received the treatment that is “optimal” for them, their PHQ-9 score at 12 weeks would have been 2.33 points lower than if they had received the treatment that is suboptimal for them. In Figure 2, the frequencies of the individual PAIs are shown. A PAI that is five or greater would mean that a substantial difference was predicted between the two treatments because 5 points on the PHQ-9 stand for a minimal clinically meaningful difference for individual change [71]. This was the case for 29% of the patients in this sample.

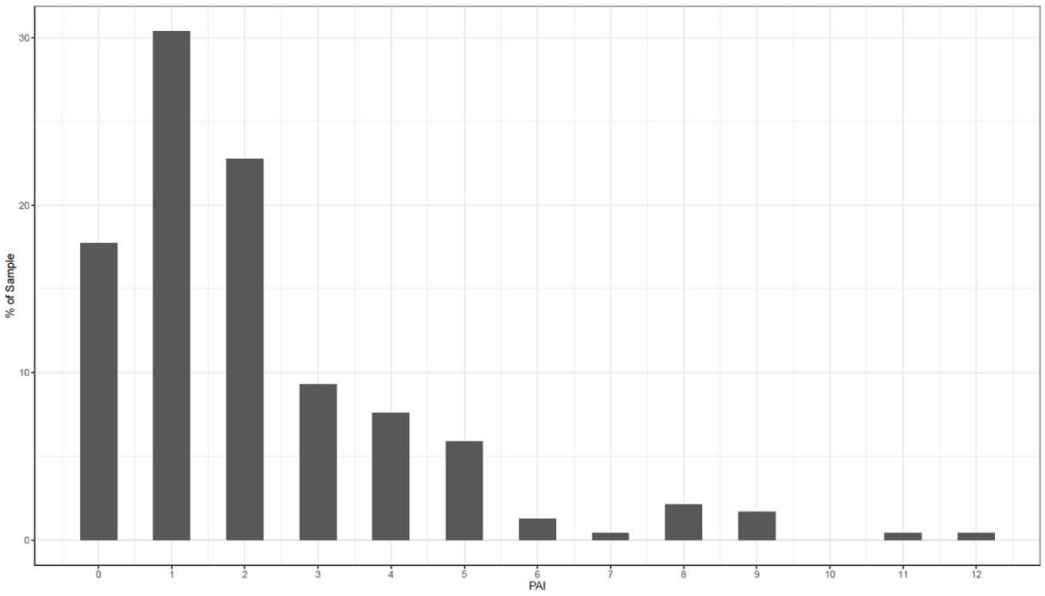

Figure 2. Frequencies of individual PAIs.

4. Discussion

Regarding the predictors of treatment outcome at 12 weeks in each of the interventions, different relevant predictors were identified for TAU and blended treatment, respectively. In the TAU condition, a lower pre-treatment PHQ-9 score, less consultations with a psychiatrist, and less days in a day-time treatment program in a psychiatric hospital, higher expectancy, receiving TAU in Denmark, and more days using outpatient psychotherapeutic services in the four months prior to the study predicted a better treatment outcome, i.e., a lower PHQ-9 score (at 12 weeks). In contrast, in the blended treatment condition, a lower pre-treatment PHQ-9 score, not being widowed, less hospital admissions and consulting self-help, a higher expectancy for improvement and a higher EQ-5D score predicted lower scores at 12 weeks, thus a better treatment outcome.

To offer an initial interpretation of our findings, the distinction between prescriptive and prognostic predictors is used. Prognostic variables predict treatment outcome regardless of treatment condition [30, 38]. In contrast, prescriptive variables may support differential indications by predicting whether a patient will benefit more from one treatment in comparison

to another. In the present study, the pre-treatment depressive symptomatology and treatment expectancy are the only *prognostic* predictors, i.e., the only variables that predict treatment outcome in both conditions. This is in line with previous research that has found pre-treatment symptomatology and expectancy to be important predictors of treatment outcome, in the sense that higher symptomatology before treatment predicts worse end-state symptomatology [30, 34, 36, 38, 72, 73] and higher expectancy for improvement predicts better treatment outcome [74, 75]. Interestingly, for internet-based treatments, higher baseline symptomatology is not necessarily a negative predictor of treatment outcome. More often the opposite is found, i.e. that higher depressive symptomatology pre-treatment predicts better treatment outcome [76-78]. This might be partly explained by the efficacy nature of previous RCTs in comparison to the routine care and effectiveness nature of the current study.

Regarding the *prescriptive* predictors, our findings partly corroborate findings from previous studies predicting treatment outcome for patients with depression. With regard to prescriptive predictors of the blended treatment condition, a lower quality of life, being widowed predicted worse treatment outcome. In contrast to the present result, the study by Huibers and colleagues [40] found higher quality of life to be a *prognostic* predictor, i.e., to predict favorable outcome irrespective of treatment condition.

For the TAU condition, more consultations with a psychiatrist, and more days in a day-time treatment program in a psychiatric hospital predicted worse treatment outcome. This could mean that patients' symptomatology and patients' general functioning is too severe to be able to profit from TAU. Furthermore, to our knowledge, there are no international studies regarding psychotherapy of depression indicating country as a relevant predictor of outcome.

Healthcare utilization within the four months prior starting treatment was found to be a prescriptive predictor in both conditions. Healthcare uptake may be a proxy for a higher somatic or mental burden and/or may represent a more severe symptomatology of depression. In

previous studies, more complex cases (e.g., with chronic symptoms and psychiatric comorbidities) or more severe depressive symptomatology predicted a worse therapy outcome [29, 38, 79-81]. Interestingly, more hospital admissions only predicted worse treatment outcome in the blended treatment condition. A possible interpretation of this finding is the fact that the online modules in the blended treatment protocol are highly standardized to target depressive symptomatology. As a result, they may have not sufficiently addressed comorbid symptoms. Somatic comorbidities in patients with depression are not a scarcity and have an influence on individual treatment response and illness course because they can complicate treatment [81]. Furthermore, number of hospital admissions may also be a proxy for case complexity and higher mental burden which in turn may have a negative impact on treatment outcome.

This study's results demonstrate that BMA makes it possible to use a limited set of baseline variables to predict treatment outcome. This is in line with a recent study by Bremer and colleagues [42] that showed the feasibility of providing personalized treatment recommendations at baseline regarding clinical and cost-effectiveness using a subsample of the current study by evaluating various machine learning techniques. Moreover, the present study showed that despite sharing the same diagnosis, patients might benefit more from different treatments. The current study found a PAI of 2.33, indicating that patients could have a PHQ-9 score at 12 weeks that is on average more than two points lower if they receive their model-determined optimal treatment in comparison to the suboptimal treatment. This value is in the range of other studies that have found PAIs ranging from 1.35 [38] up to 8.9 [40]. Importantly, for almost one third (29%) of the patients in the present study, a substantial difference was predicted between the two treatment modalities as the individual PAI was 5 or greater. This result is in line with increasing evidence suggesting that differential treatment responses are not rare and might play an important role for an individual patient and the health care system [30].

The current study has several limitations. First of all, the relatively small sample size did not allow us to build and test the models in separate samples. Using the same sample for testing and building the model might lead to a potential risk of overconfidence. Nevertheless, if studies are designed to develop and validate prescriptive prediction scores that can be tested in future hypothesis-driven confirmatory studies, a smaller sample size might be legitimate [82]. Secondly, the results are based solely on self-reports and future studies should also include observer ratings. Third, the restricted set of baseline measures is another limitation. Constructs such as personality traits or the familiarity with computers have not been assessed but may have influenced engagement with the online component. Relatedly, people with a low socioeconomic status or senior citizens may not have been well represented in the present study sample. Such groups may not have the opportunity to benefit from blended treatment because they may not have access to a computer or a smartphone and/or lack the knowledge to use them. As a consequence, the restricted sample in the present study limits the generalizability of the results. Furthermore, the current study predicts treatment outcome after 12 weeks. Future studies should predict long-term treatment outcome [83]. Finally, we have followed a data-driven approach instead of a theory-driven approach. Although the two methods should be considered complementary [84], a disadvantage of data-driven research is that it is not experiential and relying on the data alone might not capture the whole picture. Relatedly, the predictors found in the current study need to be validated and replicated in future hypothesis-driven studies.

In spite of the limitations above, the current study is promising to contribute to the further understanding of treatment for depression because it investigates implications for the use of blended treatment for patients with depression. Clinical practice should consider factors found to play a role in the treatment and processes of change to provide the optimal treatment for each individual. For example, the quality of life should be taken into account as these patients may need a more intense treatment protocol that integrates face-to-face interventions with web-

based technologies. This interpretation is in line with the notion that more severely depressed patients see the availability of an online program between face-to-face sessions as an advantage of blended treatment [85]. Furthermore, healthcare utilization should be evaluated prior to treatment selection because it can give valuable information about the patients' needs, treatment history and course of illness. In addition, the predictors found to be important in this study and in previous studies could be taken into account to make an informed treatment recommendation in clinical practice. However, future studies with larger samples and more advanced techniques are necessary to validate the current findings and the identified predictors have to be tested within clinical routine treatment settings. Moreover, prospective studies need to integrate the PAI in the diagnostics process at the beginning of a treatment.

5. Conclusions

To conclude with the first aim of the study, two *prognostic* predictors, namely pre-treatment symptomatology and treatment expectancy were found. Furthermore, several *prescriptive* predictors were found predicting treatment outcome respective of each of the two conditions. Some of our findings are in line with previous research, but other variables such as baseline healthcare utilization have not been investigated in this context. The interpretations regarding the prognostic and prescriptive predictors need to be tested empirically because they are somewhat speculative. Furthermore, this study showed an advantage of model-determined treatment allocation to TAU or blended treatment as one third of the participants had a PAI larger than 5 which means they would have improved significantly if they had received their "optimal" treatment. Although the results need to be validated in future hypothesis-driven studies, the predictors found to be important in the current study should be taken into account to make an informed treatment recommendation in clinical practice.

Conflicts of Interest: Authors report no conflict of interest.

References

1. World Health Organization. *Depression*. 2018 [cited 2019 21 November]; Available from: https://www.who.int/news-room/fact-sheets/detail/depression.
2. World Health Organization. *The global burden of disease: 2004 update*. 2008; Available from: https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/.
3. Kemmeren, L., et al., *Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial.* BMC psychiatry, 2016. **16**(113). doi: 10.1186/s12888-016-0818-5
4. Karasu, T.B., et al., *Treatment of patients with major depressive disorder.* American Psychiatric Association Practice Guidelines for the Treatment of Psychiatric Disorders Compendium, 2000: p. 413-495.
5. Cuijpers, P., *Psychotherapies for adult depression: recent developments.* Current Opinion in Psychiatry, 2015. **28**(1): p. 24-29.
6. Carlbring, P., et al., *Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis.* Cognitive Behaviour Therapy, 2018. **47**(1): p. 1-18.
7. Arnberg, F.K., et al., *Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness.* PloS One, 2014. **9**(5): p. e98118.
8. Gerhards, S., et al., *Economic evaluation of online computerised cognitive–behavioural therapy without support for depression in primary care: randomised trial.* The British Journal of Psychiatry, 2010. **196**(4): p. 310-318.

9. Richards, D. and T. Richardson, *Computer-based psychological treatments for depression: a systematic review and meta-analysis*. Clinical Psychology Review, 2012. **32**(4): p. 329-342.

10. Krieger, T., et al., *Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: design of a pragmatic randomized controlled trial.* BMC Psychiatry, 2014. **14**(1): p. 285.

11. Hickie, I.B., et al., *Practitioner-supported delivery of internet-based cognitive behaviour therapy: evaluation of the feasibility of conducting a cluster randomised trial.* Medical Journal of Australia, 2010. **192**: p. S31-S35.

12. Høifødt, R.S., et al., *The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial.* Journal of Medical Internet Research, 2013. **15**(8): p. e153.

13. Kessler, D., et al., *Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial.* The Lancet, 2009. **374**(9690): p. 628-634.

14. Kooistra, L.C., et al., *Blended vs. face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost-effectiveness trial.* BMC Psychiatry, 2014. **14**(1): p. 290.

15. Robertson, L., et al., *Using the Internet to enhance the treatment of depression.* Australasian Psychiatry, 2006. **14**(4): p. 413-417.

16. Berger, T., et al., *Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial.* Journal of Affective Disorders, 2018. **227**: p. 455-462.

17. Thase, M.E., et al., *Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT.* American Journal of Psychiatry, 2017. **175**(3): p. 242-250.

18. Zwerenz, R., et al., *Online Self-Help as an Add-On to Inpatient Psychotherapy: Efficacy of a New Blended Treatment Approach.* Psychotherapy and Psychosomatics, 2017. **86**(6): p. 341-350.

19. Erbe, D., et al., *Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review.* Journal of Medical Internet Research, 2017. **19**(9): p. e306.

20. Wright, J.H., et al., *Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time.* American Journal of Psychiatry, 2005. **162**(6): p. 1158-1164.

21. Clough, B.A. and L.M. Casey, *Technological adjuncts to increase adherence to therapy: a review.* Clinical Psychology Review, 2011. **31**(5): p. 697-710.

22. Riper, H., et al., *Preventie & eMental-Health-Prevention & eMental-Health.* Amsterdam: Vrije Universiteit commissioned by ZonMw, 2013.

23. van der Vaart, R., et al., *Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study.* BMC Psychiatry, 2014. **14**(1): p. 355.

24. Wentzel, J., et al., *Mixing online and face-to-face therapy: how to benefit from blended care in mental health care.* JMIR Mental Health, 2016. **3**(1): p. e9.

25. Ruwaard, J. and R. Kok, *Wild West eHealth: Time to hold our horses?* European Health Psychologist, 2015. **17**(1): p. 45-49.

26. Zilcha-Mano, S., *Major developments in methods addressing for whom psychotherapy may work and why.* Psychotherapy Research, 2018. **29**(6): p. 693-708.

27. Kazdin, A.E., *Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care.* American Psychologist, 2008. **63**(3): p. 146.

28. Cuijpers, P., et al., *The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size.* Psychological Medicine, 2010. **40**(2): p. 211-223.

29. Fournier, J.C., et al., *Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression.* Journal of Consulting and Clinical Psychology, 2009. **77**(4): p. 775.

30. Cohen, Z.D. and R.J. DeRubeis, *Treatment selection in depression.* Annual Review of Clinical Psychology, 2018. **14**: p. 209-236.

31. Fisher, A.J. and J.F. Boswell, *Enhancing the personalization of psychotherapy with dynamic assessment and modeling.* Assessment, 2016. **23**(4): p. 496-506.

32. Hollon, S.D., et al., *Recent Developments in the Treatment of Depression.* Behavior Therapy, 2019. **50**(2): p. 257-269.

33. Hamburg, M.A. and F.S. Collins, *The path to personalized medicine.* New England Journal of Medicine, 2010. **363**(4): p. 301-304.

34. van Bronswijk, S.C., et al., *A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder.* Depression and anxiety, 2018. **36**(3): p. 252-261.

35. Becker, D., et al., *Predictive modeling in e-mental health: A common language framework.* Internet Interventions, 2018. **12**: p. 57-67.

36. DeRubeis, R.J., et al., *The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration.* PloS One, 2014. **9**(1): p. e83875.

37. Cohen, Z.D., et al., *Recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression.* 2017.

38. Friedl, N., et al., *Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE).* Psychotherapy Research, 2019: p. 1-13.

39. DeRubeis, R.J., et al., *Cognitive therapy vs medications in the treatment of moderate to severe depression.* Archives of General Psychiatry, 2005. **62**(4): p. 409-416.

40. Huibers, M.J., et al., *Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach.* PloS One, 2015. **10**(11): p. e0140771.

41. van Breda, W., et al., *Predicting therapy success for treatment as usual and blended treatment in the domain of depression.* Internet interventions, 2018. **12**: p. 100-104.

42. Bremer, V., et al., *Predicting therapy success and costs for personalized treatment recommendations using baseline characteristics: Data-driven analysis.* Journal of Medical Internet Research, 2018. **20**(8): p. e10275.

43. Kleiboer, A., et al., *European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries.* Trials, 2016. **17**(1): p. 387.

44. Sheehan, D.V., et al., *The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10.* The Journal of Clinical Psychiatry, 1998.

45. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* Journal of General Internal Medicine, 2001. **16**.

46. Meyer, B., et al., *Effectiveness of a Novel Integrative Online Treatment for Depression (Deprexis): Randomized Controlled Trial.* Journal of Medical Internet Research, 2009. **11**(2): p. e15.

47. Mathiasen, K., et al., *Blended CBT versus face-to-face CBT: a randomised non-inferiority trial.* BMC Psychiatry, 2016. **16**(1): p. 432.

48. Kroenke, K., *Enhancing the clinical utility of depression screening.* CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 2012. **184**(3): p. 281-282.

49. Kroenke, K. and R.L. Spitzer, *The PHQ-9: a new depression diagnostic and severity measure.* Psychiatric Annals, 2002. **32**(9): p. 509-515.

50. Manea, L., S. Gilbody, and D. McMillan, *Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis.* Canadian Medical Association Journal, 2012. **184**(3): p. E191-E196.

51. EuroQol, *EuroQol - a new facility for the measurement of health-related quality of life.* Health Policy, 1990. **16**(3): p. 199-208.

52. Bouwmans, C., et al., *Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P).* BMC Health Services Research, 2013. **13**(1): p. 217.

53. Devilly, G.J. and T.D. Borkovec, *Psychometric properties of the credibility/expectancy questionnaire.* Journal of Behavior Therapy and Experimental Psychiatry, 2000. **31**(2): p. 73-86.

54. Stekhoven, D.J., *missForest: Nonparametric Missing Value Imputation using Random Forest. R package version, 1, 4.* 2013.

55. Stekhoven, D.J. and P. Buhlmann, *MissForest--non-parametric missing value imputation for mixed-type data.* Bioinformatics, 2012. **28**(1): p. 112-8.

56. Waljee, A.K., et al., *Comparison of imputation methods for missing laboratory data in medicine.* BMJ Open, 2013. **3**(8): p. e002847.

57. Peduzzi, P., et al., *Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates.* Journal of Clinical Epidemiology, 1995. **48**(12): p. 1503-1510.

58. Vittinghoff, E. and C.E. McCulloch, *Relaxing the rule of ten events per variable in logistic and Cox regression.* American Journal of Epidemiology, 2007. **165**(6): p. 710-718.

59. Raftery, A.E., D. Madigan, and J.A. Hoeting, *Bayesian model averaging for linear regression models.* Journal of the American Statistical Association, 1997. **92**(437): p. 179-191.

60. Stock, E.M., et al., *A Bayesian model averaging approach to examining changes in quality of life among returning Iraq and Afghanistan Veterans.* International Journal of Methods in Psychiatric Research, 2014. **23**(3): p. 345-358.

61. Babyak, M.A., *What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models.* Psychosomatic Medicine, 2004. **66**(3): p. 411-421.

62. Fragoso, T.M., W. Bertoli, and F. Louzada, *Bayesian model averaging: A systematic review and conceptual classification.* International Statistical Review, 2018. **86**(1): p. 1-28.

63. Fernandez, C., E. Ley, and M.F. Steel, *Benchmark priors for Bayesian model averaging.* Journal of Econometrics, 2001. **100**(2): p. 381-427.

64. Fernandez, C., E. Ley, and M.F. Steel, *Model uncertainty in cross-country growth regressions.* Journal of applied Econometrics, 2001. **16**(5): p. 563-576.

65. Ley, E. and M.F.J. Steel, *On the effect of prior assumptions in Bayesian model averaging with applications to growth regression.* Journal of Applied Econometrics, 2009. **24**(4): p. 651-674.

66. Steel, M.F., *Bayesian model averaging and forecasting.* Bulletin of EU and US Inflation and Macroeconomic Analysis, 2011. **200**: p. 30-41.

67. Clyde, M.A. *BAS: Bayesian Variable Selection and Model Averaging using Bayesian Adaptive Sampling. R package version 1.5.3*. 2018; Available from: https://cran.r-project.org/web/packages/BAS/BAS.pdf.

68. Clyde, M.A., J. Ghosh, and M.L. Littman, *Bayesian adaptive sampling for variable selection and model averaging.* Journal of Computational and Graphical Statistics, 2011. **20**(1): p. 80-101.

69. Efron, B. and G. Gong, *A leisurely look at the bootstrap, the jackknife, and cross-validation.* The American Statistician, 1983. **37**(1): p. 36-48.

70. Deisenhofer, A.K., et al., *Individual treatment selection for patients with posttraumatic stress disorder.* Depression and Anxiety, 2018. **35**(6): p. 541-550.

71. Lowe, B., et al., *Monitoring depression treatment outcomes with the patient health questionnaire-9.* Medical Care, 2004. **42**(12): p. 1194-201.

72. Dinga, R., et al., *Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach.* Translational Psychiatry, 2018. **8**(1): p. 241.

73. Hamilton, K.E. and K.S. Dobson, *Cognitive therapy of depression: pretreatment patient predictors of outcome.* Clinical Psychology Review, 2002. **22**(6): p. 875-893.

74. Constantino, M.J., et al., *Expectations.* Journal of Clinical Psychology, 2011. **67**(2): p. 184-192.

75. Greenberg, R.P., M.J. Constantino, and N. Bruce, *Are patient expectations still relevant for psychotherapy process and outcome?* Clinical Psychology Review, 2006. **26**(6): p. 657-678.

76. Spek, V., et al., *Predictors of outcome of group and internet-based cognitive behavior therapy.* Journal of Affective Disorders, 2008. **105**(1): p. 137-145.

77. Andersson, G. and E. Hedman, *Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings.* Verhaltenstherapie, 2013. **23**(3): p. 140-148.

78. Bower, P., et al., *Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data.* BMJ (Clinical research ed.), 2013. **346**: p. f540-f540.

79. Thase, M.E., et al., *Response to cognitive-behavioral therapy in chronic depression.* Journal of Psychotherapy Practice & Research, 1994. **3**(3): p. 304-314.

80. Blom, M.B., et al., *Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression.* Journal of Affective Disorders, 2007. **104**(1-3): p. 119-126.

81. Delgadillo, J., et al., *Case complexity as a guide for psychological treatment selection.* Journal of Consulting and Clinical Psychology, 2017. **85**(9): p. 835-853.

82. Luedtke, A., E. Sadikova, and R.C. Kessler, *Sample Size Requirements for Multivariate Models to Predict Between-Patient Differences in Best Treatments of Major Depressive Disorder.* Clinical Psychological Science, 2019. **7**(3): p. 445-461.

83. van Bronswijk, S.C., et al., *Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: cognitive therapy or interpersonal psychotherapy?* Psychological Medicine, 2019: p. 1-11.

84. Gillan, C.M. and R. Whelan, *What big data can do for treatment in psychiatry.* Current Opinion in Behavioral Sciences, 2017. **18**: p. 34-42.

85. Urech, A., et al., *A patient post hoc perspective on advantages and disadvantages of blended cognitive behaviour therapy for depression: A qualitative content analysis.* Psychotherapy Research, 2018: p. 1-13.

3 Discussion

The current situation of global mental health suggests that mental health issues will continue to play a big role in society and the consequences will increase with time. The number of people in need of treatment will grow simultaneously with the number of patients who do not benefit from first-line treatment as expected. As research has shown, patients vary in their illness course and their treatment response. This highlights the need for personalized treatments (Cuijpers & Christensen, 2017). In the last years, new advances in psychotherapy research and precision medicine have shown promising results for treatment selection. With the use of multivariable prediction methods, the identification of prescriptive predictors for treatment response is within reach. This doctoral thesis gives an overview of the development of process-outcome research and its advances over time. While the Generic Model of Psychotherapy is introduced as a theoretical framework to identify variables that have a relationship with the psychotherapy outcome, the two studies presented in this thesis use the PAI approach to identify prescriptive predictors for different treatments for depression. On the one hand, the studies show the feasibility of the PAI approach and on the other hand, they shed a light on new treatment approaches like assimilative integration and blended treatment for patients with depression. The first study compares CBT to CBT-EE which is a CBT that integrates emotion focused principles. Important predictors for each treatment condition are identified and model-determined treatment allocation is compared to randomization (see Chapter 2.1). The second study uses the same methodology to identify predictors for treatment outcome in either TAU or blended treatment and compares model-determined treatment allocation to randomization (see Chapter 2.2). In the following, the results of the two studies will be discussed in more detail, their limitations will be critically reviewed, and future implications of process-outcome research and treatment selection will be discussed.

3. 1 Determining optimal treatment allocation to CBT vs CBT-EE

Study 1 identified the most important predictors determining optimal treatment allocation to the integrative CBT-EE or standard CBT. In a second step, it investigated if model-determined treatment allocation using predictive information results in a better treatment outcome for patients with depression. Pre-treatment depressive symptomatology was the only *prognostic* predictor found, thus the only variable that predicted treatment outcome in both treatment conditions in the sense that lower pre-treatment depressive symptomatology predicts a better treatment outcome, i.e. lower post-treatment symptomatology. Regarding the *prescriptive* predictors, different relevant variables in both conditions were found. For the CBT-EE condition, lower age, not being separated or divorced, having accomplished a higher education than an apprenticeship, no comorbid anxiety disorder, no comorbid axis-II disorder, lower psychopathology, lower self-focused rumination, and more hope for improvement predicted a lower post-treatment BDI-II score. In the CBT condition, female gender, fewer previous depressive episodes, no recurrent depression and a lower current incongruence regarding avoidance motives predicted a better treatment outcome.

Previous research has found a variety of different predictors due to the heterogeneity in research methods and different sets of baseline measures. But the prognostic predictor found in this study is in line with multiple earlier studies who have found pre-treatment symptomatology to be a prognostic predictor as well. To be exact, previous studies have shown that higher pre-treatment severity is related to higher post-treatment symptomatology that is defined as worse treatment outcome (Cohen & DeRubeis, 2018; DeRubeis, Cohen, et al., 2014; Dinga et al., 2018; Hamilton & Dobson, 2002; van Bronswijk et al., 2018). As mentioned in Chapter 2.1.3, baseline measures that have an emotional focus have been included as possible predictors. Interestingly, variables that target constructs like emotion

regulation skills, negative mood regulation, ambivalence over the expression of emotion, or self-compassion did not seem to be important to predict treatment outcome in CBT-EE. Therapist responsiveness could be held accountable for this in the sense that therapists responded differentially to the emotional baseline characteristics that patients displayed regardless of treatment condition which then prevented potential differential predictions (Stiles, Honos-Webb, & Surko, 1998). Another explanation could be that both treatment conditions can deal equally well with emotional variables meaning that pre-treatment characteristics in fact did not have any differential predictive power. With regard to the prescriptive predictors of the CBT condition, a study that compared CBT to IPT found female gender to be a prognostic predictor of lower depressive symptomatology at post-treatment (Huibers et al., 2015). In addition, previous research has found previous depressive episodes and recurrent depression to be a prognostic predictor (Blom et al., 2007; Fournier et al., 2009; Thase, Reynolds, Frank, & Simons, 1994) whereas it has only predicted treatment outcome for CBT in the current study. Moreover, it is surprising that some of the prescriptive predictors of the CBT-EE condition, such as age, educational level, and employment have been found to be either prognostic or prescriptive in studies that compared CBT, CT or treatment with antidepressant medication (Chekroud et al., 2016; DeRubeis, Cohen, et al., 2014; Fournier et al., 2009).

Furthermore, we found that self-focused rumination decreases therapy outcome specifically in CBT-EE. As emotion-focused interventions have the goal to change maladaptive self-associations, a ruminative self-processing style might make it hard to engage in interventions like the two-chair or empty chair exercises (Greenberg, 2002). Hope for improvement was also an important predictor in the CBT-EE condition. It is possible that more hope for improvement helps the patient “endure” burdensome emotions that are

activated with emotion-focused techniques because of a possible better outcome in the long run (Westermann, grosse Holtforth, & Michalak, 2019, in press).

We have included motivational factors in the baseline measures, and in fact, lower avoidance incongruence (making less aversive experiences) has found to be an important predictor in the CBT condition. Although motivational factors have not been examined in earlier research, this finding highlights the relevance to do so. Having stronger aversive experiences in life at intake decreases the chance for good therapy outcomes. This is a factor that can be crucial in treatment selection and should be acknowledged and further investigated.

Study 1 shows the feasibility of using the PAI approach to compare two forms of CBT, the standard form and an assimilative integration (CBT-EE) which has not been done up to date. An advantage of applying PAI research to the field of assimilative integration is that the therapists can conduct each of the compared interventions while remaining in his or her theoretical framework. The empirical evidence supporting the efficacy of assimilative integration is growing and it is common in clinical practice that therapists stay within the theoretical framework that they are trained in but use techniques of different approaches that potentially utilize other change mechanisms (Castonguay, Eubanks, Goldfried, Muran, & Lutz, 2015; Grosse Holtforth et al., 2019; Lampropoulos, 2001). This is in line with Grawe's *General Psychotherapy* which can be seen as a model of continuous expansion of a therapist's original approach. For Grawe, *General Psychotherapy* is a guiding principle for clinicians with the aim to use all well-established therapeutic interventions in order to achieve the best possible treatment result for each individual patient (Grawe, 1999).

To our knowledge, this is the first study that uses BMA in combination with the PAI. We found a PAI of 1.35 indicating that patients could have had a post-treatment BDI-II score

that is more than one point lower if they would have received their optimal treatment in comparison to their suboptimal treatment. Other studies have found PAIs ranging from 1.6 (Cohen et al., 2019) up to 8.9 (Huibers et al., 2015). This can be explained by the nature of the two treatments compared. As mentioned before, CBT and CBT-EE share the same theoretical background and might not differ as much as the treatments that have been compared in earlier studies. The *same therapist* design of the RCT could have lowered the PAI as well. Regardless of the small value of the PAI, for 46% of the patients in this sample, a substantial difference was predicted between the two treatments as the PAI was 5 or greater.

3.2 Treatment selection in TAU vs blended treatment

Using the same approach as in Study 1, Study 2 identified the most important predictors that determine optimal treatment allocation to TAU or blended treatment. Moreover, model-determined treatment allocation that uses this predictive information was compared to randomization. Before we have a look at the prescriptive predictors that predict treatment outcome differentially, the two *prognostic* predictors that were found will be discussed. Pre-treatment depressive symptomatology and treatment expectancy were the only variables that predicted PHQ-9 scores at 12 weeks. This finding is partly in line with study 1 and previous research where higher symptomatology before treatment has often been found to predict worse outcome in the sense of higher post-treatment symptomatology whereas higher expectancy for improvement predicts better outcome, thus lower post-treatment symptomatology (Cohen & DeRubeis, 2018; DeRubeis, Cohen, et al., 2014; Dinga et al., 2018; Friedl, Berger, Krieger, Caspar, & Grosse Holtforth, 2019; Hamilton & Dobson, 2002; van Bronswijk et al., 2018). For internet-based treatments, previous research suggests that higher pre-treatment symptomatology is often a predictor for higher improvement, thus higher change scores (Bower et al., 2013; Spek, Nyklíček, Cuijpers, & Pop, 2008). As most of the

studies that have been conducted in previous years are RCTs, this difference might be explained by the routine care and effectiveness nature of the E-COMPARED study.

In the blended treatment condition, a lower quality of life and being widowed predicted worse treatment outcome at 12 weeks. Whereas quality of life is a prescriptive predictor here, it has predicted favorable outcome irrespective of treatment condition, thus it has acted as a prognostic predictor in previous studies (Huibers et al., 2015). Furthermore, healthcare utilization within the four months before starting treatment in the sense of more hospital admissions predicted worse treatment outcome. Healthcare uptake may be an indicator of a higher somatic or mental burden and may represent a more severe depressive symptomatology. This would be in line with earlier research that has found more complex cases (e.g. with psychiatric comorbidities and chronic symptoms) and more severe depressive symptomatology to predict worse treatment outcome (Blom et al., 2007; Delgadillo, Huey, Bennett, & McMillan, 2017; Fournier et al., 2009; Friedl et al., 2019; Thase et al., 1994). As this was only found for the blended treatment, it is possible that the highly standardized online modules did not sufficiently address comorbid symptoms. For the TAU condition, more consultations with a psychiatrist, and more days in a day-time treatment program in a psychiatric hospital predicted worse treatment outcome at 12 weeks. This means that patient's healthcare utilization prior to treatment was higher which can speak for a poorer general functioning of the patients. As this is a predictor in the TAU condition only, it is possible that patient's functioning was not sufficient to benefit from TAU.

Study 2 found a PAI of 2.33 indicating that patients could have a PHQ-9 score that is on average two points lower if they receive their model-determined optimal treatment as opposed to receiving the suboptimal treatment. The results demonstrate the feasibility of BMA to use a restricted set of baseline variables to predict treatment outcome. Even though different forms of treatment have been compared before (Cohen et al., 2019; Friedl et al.,

2019; Huibers et al., 2015), comparing a face-to-face therapy with blended treatment is a new approach. This could be highly useful in increasing treatment response rates in patients with depression because the efficacy and feasibility of blended treatments have been shown (Erbe, Eichert, Riper, & Ebert, 2017; Kooistra et al., 2016). Blended treatments can offer a solution for the problem of accessibility and affordability of treatment which has been mentioned in Chapter 1 while it has the ability to promote patient participation and reduce direct medical costs (Kenter et al., 2015; Kooistra et al., 2016; Kooistra et al., 2014). Although the PAI found here lies within the range of PAIs found in other studies, e.g. 1.35 (Friedl et al., 2019) up to 8.9 (Huibers et al., 2015), it is questionable whether they can be compared because of the different treatment modalities.

3.3 Limitations of the studies presented

Overall, the findings of the studies presented are promising. The results may contribute to the further understanding of treatment for depression. Nevertheless, there are certain difficulties and limitations in the field of precision medicine and treatment selection. In the following, the most important limitations which are generally associated with the research presented here, are discussed.

A common problem in research that plays an essential role in precision medicine and treatment selection is small sample sizes. Ideally, the sample size would be big enough to use a subsample to build the prediction models and to subsequently use another subsample to test and validate the models. The sample size of the two studies presented here did not allow us to train and test the models in different subsamples which leads to a potential risk of overconfidence. A recent study analyzed the sample size requirements for multivariate prescriptive prediction models that guide treatment selection for patients with depression. Results suggest that at least 300 patients per treatment arm are required to have sufficient

statistical power. Although this is a considerably higher number than the sample sizes in most studies, it is important to mention that this suggestion is only applicable to populations whose parameters are in the range of the ones used in the study that generated the results. Furthermore, smaller sample sizes can be justified if studies are designed to develop and validate prescriptive prediction scores that can then be tested in hypothesis-driven confirmatory studies. As an example, this is the case for the two studies presented here (Luedtke et al., 2019; Petkova et al., 2017). Another problem common in psychotherapy research is the generalizability. When building prediction models, we do not know how the model would perform if applied to populations that have baseline scores outside the range of the populations that were used for building the models. Validation samples are needed to solve this issue but as mentioned above, large datasets are required for that.

Moreover, it is a fundamental problem that so far hardly any consistent predictors have been found throughout the years. One possible explanation for this might be therapist responsiveness which is defined as a therapist's behavior that is influenced by emerging context and in particular by patient's behavior and changing characteristics. When therapists respond to patient's requirements throughout the therapy process, they defeat research designs of process-outcome research by performing unpredictable and adapting behaviors (Kramer & Stiles, 2015). This is why responsiveness makes it impossible to draw causal conclusions in research and it undermines conclusions that are drawn based on linear statistics and linear reasoning (Kramer & Stiles, 2015; Stiles, 2009; Stiles et al., 1998). Interestingly, DeRubeis and colleagues found that the correlation between the measurement of therapy quality and outcome is quite small under conditions that are common in psychotherapy research. It appears that some patients improve regardless of the therapy quality, and others will not improve no matter how good or bad the therapy quality is. As a result, they identified various patient response patterns namely the spontaneous remitter, the easy patient, the pliant patient,

the challenging patient and the intractable patient (DeRubeis, Gelfand, German, Fournier, & Forand, 2014). Another explanation may be that research on treatment selection and precision medicine has exclusively focused on pre-treatment variables although it may be essential to include and investigate the impact of early process variables. Only when a therapy has started may patterns appear which predict a relatively consistent positive or negative course. This aspect is given more attention in the Generic Model of Psychotherapy and in process-outcome research in general. As laid out in chapter 1.2., the "process" is the heart of the Generic Model. It is the part that contains early process variables that potentially influence therapy outcomes.

Finally, there is a lack of comparability of the different studies in the field of treatment selection. Two reasons can be held accountable for this: (1) different methods and approaches to prescriptive modeling are used which result in different outcome measures that may be hard to compare and (2) studies use different sets of baseline measures. This leads to a great variety of predictors that are found and makes it difficult for clinicians to know what predictors to rely on when making a treatment recommendation.

3.4 The future of treatment selection and precision medicine

The era of treatment selection and the use of prediction algorithms in psychotherapy research is still relatively young. With the increase of research done in this field, the necessity, as well as the advantages of personalized treatments become more salient than ever. We need to limit the number of patients who do not respond to first-line treatments and we need to reduce the average time that patients need to recover. If we can manage to provide to each patient the treatment that is likely to yield the best result for him or her, we can not only reduce the suffering from symptoms of depression but we can ultimately impact the economic productivity of each individual (Laynard, Clark, Knapp, & Mayraz, 2007). As mentioned in

the previous chapters, accessibility and availability of psychotherapy and pharmacotherapy are not always given which is why the limited or costly resources should be well allocated in order to result in cost- and time- effectiveness (Cohen & DeRubeis, 2018).

Another point that needs attention is the methodology that is used to advance the clinical utility of treatment selection. Early on, approaches like STS or the Generic Model of Psychotherapy have taken into account multiple variables that may influence the treatment outcome. Precision medicine focuses on fine-grained individual differences because of the great importance of variability in treatment outcomes. What we need is actionable, prescriptive information about which patients benefit most from what treatment (Zilcha-Mano, 2018). This is only possible with the use of multivariable treatment selection approaches. Instead of relying on one single moderator, we need to take into consideration the overwhelming amount of therapist, patient and setting factors. New statistical methodologies like machine learning make this endeavor possible (Cohen & DeRubeis, 2018).

The present doctoral thesis and the two included studies show several examples of multivariable treatment selection and its feasibility. So far, they have failed to have a meaningful impact on client care because the identified predictors need to be tested prospectively within clinical routine treatment settings. Future studies should evaluate the feasibility and the added value of integrating a treatment selection model during the diagnostic process at the beginning of psychotherapy. Furthermore, for clinicians to take into account the important predictors found in previous studies to make an informed treatment recommendation for their patients, future studies with larger samples are necessary to validate them first. Another problem of the PAI, other older (e.g. STS) but also newer approaches (e.g. machine learning) is that they are atheoretical and have not yet promoted theoretical understanding of prescriptive and process-outcome research. It would be important that the new approaches are also theoretically grounded or promote theory development. This is where

the Generic Model of Psychotherapy can be helpful. To conclude, not only new methodologies are useful when optimizing patient care because newer, updated versions of older constructs can be handy, as well.

3.4.1 New Generic Model of Psychotherapy

The Generic Model of Psychotherapy already exists for over 25 years. It has not only acted as a basis for integrating research findings, it has also guided research to deconstruct, demystify and integrate clinical practice theories of psychotherapy. That said, it goes beyond summarizing findings of particular studies because it can expand to include new aspects of the therapeutic process that may be important (Orlinsky, 2009; Orlinsky et al., 2004). The Generic Model of Psychotherapy has also been used to guide exploratory psychotherapy research that aims to identify emerging patterns of relationships between process and outcome (Kolden, 1991).

With the advances in elaborate statistical methodologies like machine learning and the need for personalized treatment selection, one can ask *"What role can the Generic Model of Psychotherapy play in treatment selection? Will it be able to keep up with the new emerging methods?"*. In fact, there are some constructs that are not displayed in the model that might be of importance. As an example, the Generic Model only depicts linear relationships between process and outcome and phenomena like sudden gains (Tang & DeRubeis, 1999) are not taken into consideration. Moreover, several constructs that have not gotten much attention when the model was constructed and adapted or that have not been researched extensively, might have been found to play a role nowadays. Another important point to mention is that only very few meta-analyses have been included in the model (Orlinsky et al., 2004). Of course, the reason for that is the lack of meta-analyses at that time but nevertheless, they could shed a light on important relationships of processes and outcomes. Variables that have

been investigated more in the last years are for example therapist repair of alliance ruptures (Eubanks, Muran, & Safran, 2018) or premature vs completed termination (Orlinsky et al., 2004; Swift & Greenberg, 2012).

3.4.2 The future of precision mental health

To impact client care positively, it is essential to transfer the knowledge gained in research to clinical practice. The highest goal should be to translate research findings into clinical support tools that therapists can use to make an informed treatment selection. An important step in this process is the collaboration of patients and therapists. The needs and preferences of clinicians need to be matched with the needs and preferences of patients. Furthermore, pre-treatment assessments should include larger sets of variables, even though this is not always possible in RCTs. Large databases are available that have collected data electronically and could be used more frequently for treatment selection research (Perlis et al., 2012). More importantly, not only pre-treatment variables but also potential early process predictors should be taken into account like it is the case in the Generic Model of Psychotherapy. Clinicians need support tools that not only support them with selective indication but also with adaptive indication. More recently, person-specific dynamic assessments in personalized treatment selection and precision medicine have shown promising results although research is still rather scarce. With the use of dynamic assessments, it is possible to determine person-specific syndrome structures that provide unique information about psychopathology within an individual (Fernandez, Fisher, & Chi, 2017; Fisher, 2015; Fisher & Boswell, 2016). Moreover, more data should be made available to researchers around the world in the sense of open source because, in the long run, patients would benefit from sharing resources and results. Last but not least, clinical trials that compare treatment selection to alternative methods of allocation need to be designed in order

to validate results prospectively against standard allocation schemes (Cohen & DeRubeis, 2018).

3.4.3 Implementation of treatment selection in clinical practice

The last essential step for precision mental health to have an impact on client care is to test personalized psychotherapy prediction and adaptation tools in clinical practice (Cohen & DeRubeis, 2018). This is what the Trier Treatment Navigator (TTN) has been developed for. The TTN was developed based on a sample of 1234 patients with affective disorders that have been treated in the outpatient clinic of the University of Trier (Lutz, Clausen, & Deisenhofer, 2019; Lutz, Zimmermann, Muller, Deisenhofer, & Rubel, 2017). The empirical basis for the personalized treatment predictions of the TTN is provided by psychometric information that is collected from each patient at the beginning of therapy, i.e. the baseline measures. Based on this information, similar patients are first selected from the total database of already treated patients for a new patient to be treated using the next-neighbor procedure. Therapists then receive an assessment of which strategy has been successful with similar patients treated before. In addition, therapists receive an overview of potential risk areas the patient might have (Lutz et al., 2019). Several therapy process measures provide the basis for this adaptive indication. If the observed therapy course deviates from the mean therapy course of the 30 most similar patients and if the process is crossing the border to a risk area, a warning is given that the patient is currently "Not On Track". Besides, clinical support tools are displayed, which are related to the respective problems of the patient (e.g. risk/suicidal tendencies, motivation/therapy goals, therapeutic relationship, social support and emotion regulation). These support tools include treatment recommendations, clinical exercises and sample videos for the respective problem area. An RCT that evaluates the effects of the TTN is currently being conducted (Lutz et al., 2017).

The TTN is an important example of how the transfer of research results into daily practice can look like. Although we should not jump to conclusions, it may provide an opportunity for clinicians to integrate research on personalized treatment selection into their clinical work. Both researchers and clinicians have the same goal, namely, to offer patients the best possible treatment and to optimize mental health care for people in need of psychosocial support. Much work has yet to be done but if the field of precision medicine and treatment selection continues to grow as it has in recent years, we have the chance of improving a lot of people's lives worldwide.

4 References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (revised 4th ed.)*. Washington, DC: Author.

Backenstrass, M., Pfeiffer, N., Schwarz, T., Catanzaro, S. J., & Mearns, J. (2008). Reliabilität und Validität der deutschsprachigen Version der Generalized Expectancies for Negative Mood Regulation (NMR) Scale [Reliability and validity of the German version of the generalized expectancies for Negative Mood Regulation (NMR) Scale]. *Diagnostica, 54*(1), 43-51. doi:10.1026/0012-1924.54.1.43

Barber, J. P., & Muenz, L. R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: Empirical findings from the Treatment for Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*(5), 951-958. doi:10.1037//0022-006x.64.5.951

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. doi:10.1037/0022-3514.51.6.1173

Baumeister, H., & Gordon, P. (2012). Meta-review of depressive subtyping models. *Journal of Affective Disorders, 139*(2), 126-140. doi:10.1016/j.jad.2011.07.015

Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (1st ed.). New York: John Wiley & Sons.

Berking, M., & Znoj, H. (2008). Entwicklung und Validierung eines Fragebogens zur standardisierten Selbsteinschätzung emotionaler Kompetenzen (SEK-27) [Development and validation of a self-report measure for the assessment of Emotion

Regulation Skills (SEK-27)]. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 56*(2), 141-153. doi:10.23668/psycharchives.387

Beutler, L. E., & Clarkin, J. F. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions.* . New York: Brunner/Mazel.

Beutler, L. E., Clarkin, J. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient*. New York: Oxford University Press.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive psychotherapy: A practical guide to systematic treatment selection*. New York: Oxford University Press.

Beutler, L. E., & Harwood, T. M. (2002). What is and can be attributed to the therapeutic relationship? *Journal of Contemporary Psychotherapy, 32*(1), 25-33. doi:10.1023/A:1015579111666

Beutler, L. E., Harwood, T. M., Bertoni, M., & Thomann, J. (2006). Systematic Treatment Selection and Prescriptive Therapy. In G. Stricker & J. Gold (Eds.), *A casebook of psychotherapy integration* (pp. 29-41). Washington, DC: American Psychological Association.

Bleich, J., Kapelner, A., George, E. I., & Jensen, S. T. (2014). Variable selection for BART: an application to gene regulation. *The Annals of Applied Statistics, 8*(3), 1750-1781. doi:10.1214/14-AOAS755

Blom, M. B., Spinhoven, P., Hoffman, T., Jonker, K., Hoencamp, E., Haffmans, P. J., & van Dyck, R. (2007). Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression. *Journal of Affective Disorders, 104*(1-3), 119-126. doi:10.1016/j.jad.2007.03.010

Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the Gold Standard - Lessons from the History of RCTs. *New England Journal of Medicine, 374*(22), 2175-2181. doi:10.1056/NEJMms1604593

Bouwmans, C., De Jong, K., Timman, R., Zijlstra-Vlasveld, M., Van der Feltz-Cornelis, C., Tan, S. S., & Hakkaart-van Roijen, L. (2013). Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). *BMC Health Services Research, 13*(1), 217. doi:10.1186/1472-6963-13-217

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., . . . Liu, E. T.-H. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ (Clinical research ed.), 346*, f540-f540. doi:10.1136/bmj.f540

Brähler, E., Mühlan, H., Albani, C., & Schmidt, S. (2007). Teststatistische Prüfung und Normierung der deutschen Versionen des EUROHIS-QOL Lebensqualität-Index und des WHO-5 Wohlbefindens-Index [Testing and standardization of the German version of the EUROHIS-QOL and WHO-5 quality-of life-indices]. *Diagnostica, 53*(2), 83-96. doi:10.1026/0012-1924.53.2.83

Bremer, V., Becker, D., Kolovos, S., Funk, B., Van Breda, W., Hoogendoorn, M., & Riper, H. (2018). Predicting therapy success and costs for personalized treatment recommendations using baseline characteristics: Data-driven analysis. *Journal of Medical Internet Research, 20*(8), e10275. doi:10.2196/10275

Caspar, F. (1997). What Goes on in a Psychotherapist's Mind? *Psychotherapy Research, 7*(2), 105-125. doi:10.1080/10503309712331331913

Castonguay, L. G., Eubanks, C. F., Goldfried, M. R., Muran, J. C., & Lutz, W. (2015). Research on psychotherapy integration: Building on the past, looking to the future. *Psychotherapy Research, 25*(3), 365-382. doi:10.1080/10503307.2015.1014010

Cavanagh, K. (2014). Geographic inequity in the availability of cognitive behavioural therapy in England and Wales: a 10-year update. *Behavioural and Cognitive Psychotherapy, 42*(4), 497-501. doi:10.1017/S1352465813000568

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., . . . Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, *3*(3), 243-250. doi:10.1016/S2215-0366(15)00471-X

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology*, *14*, 209-236. doi:10.1146/annurev-clinpsy-050817-084746

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2019). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research, 30*(2), 1-14. doi:10.1080/10503307.2018.1563312

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry, 58*(7), 376-385. doi:10.1177/070674371305800702

Cuijpers, P., & Christensen, H. (2017). Are personalised treatments of adult depression finally within reach? *Epidemiological Psychiatric Sciences*, *26*(1), 40-42. doi:10.1017/S204579601600007X

Cuijpers, P., Van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine, 40*(2), 211-223. doi:10.1017/S0033291709006114

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674. doi:10.1126/science.2648573

Deighton, R. M., & Traue, H. C. (2006). Emotionale Ambivalenz, Körperbeschwerden, Depressivität und soziale Interaktion: Untersuchungen zur deutschen Version des "Ambivalence over Emotional Expressiveness Questionnaire "(AEQ-G18) [Emotional ambivalence, illness behavior, depression, and social interaction: Investigations for the German version of the "Ambivalence over Emotional Expressiveness Questionnaire" (AEQ-G18)]. *Zeitschrift für Gesundheitspsychologie, 14*(4), 158-170. doi:10.1026/0943-8149.14.4.158

Deisenhofer, A. K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety, 35*(6), 541-550. doi:10.1002/da.22755

Delgadillo, J., & Gonzalez Salas Duhne, P. (2019). Targeted prescription of cognitive behavioral therapy vs. person-centered counseling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology, 88*(1), 14-24. doi:10.1037/ccp0000476

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology, 85*(9), 835-853. doi:10.1037/ccp0000231

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PloS One, 9*(1), e83875. doi:10.1371/journal.pone.0083875

DeRubeis, R. J., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014). Understanding processes of change: How some patients reveal more than others-and

some groups of therapists less-about what matters in psychotherapy. *Psychotherapy Research, 24*(3), 419-428. doi:10.1080/10503307.2013.838654

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., . . . Brown, L. L. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62*(4), 409-416. doi:10.1001/archpsyc.62.4.409

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 31*(2), 73-86. doi:10.1016/S0005-7916(00)00012-4

Dinga, R., Marquand, A. F., Veltman, D. J., Beekman, A. T., Schoevers, R. A., van Hemert, A. M., . . . Schmaal, L. (2018). Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Translational Psychiatry, 8*(1), 241. doi:10.1038/s41398-018-0289-1

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician, 37*(1), 36-48. doi:10.2307/2685844

Erbe, D., Eichert, H.-C., Riper, H., & Ebert, D. D. (2017). Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. *Journal of Medical Internet Research, 19*(9), e306. doi:10.2196/jmir.6588.

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy, 55*(4), 508-519. doi:10.1037/pst0000185

EuroQol. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199-208. doi:10.1016/0168-8510(90)90421-9

Fernandez, K. C., Fisher, A. J., & Chi, C. (2017). Development and initial implementation of the Dynamic Assessment Treatment Algorithm (DATA). *PloS One, 12*(6), e0178806. doi:10.1371/journal.pone.0178806

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology, 83*(4), 825. doi:10.1037/ccp0000026

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment, 23*(4), 496-506. doi:10.1177/1073191116638735

Flückiger, C., Del Re, A., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(5), 316-340. doi:10.1037/pst0000172

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting and Clinical Psychology, 77*(4), 775. doi:10.1037/a0015401

Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review, 86*(1), 1-28. doi:10.1111/insr.12243

Friedl, N., Berger, T., Krieger, T., Caspar, F., & Grosse Holtforth, M. (2019). Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE). *Psychotherapy Research*, 1-13. doi:10.1080/10503307.2019.1664782

Grawe, K. (1976). *Differentielle Psychotherapie. Indikation und spezifische Wirkung von Verhaltenstherapie und Gesprächspsychotherapie [Differential psychotherapy. Indication and specific effects of behavior therapy and client-centered therapy]*. Bern: Hans Huber.

Grawe, K. (1997). Research-Informed Psychotherapy. *Psychotherapy Research, 7*(1), 1-19. doi:10.1080/10503309712331331843

Grawe, K. (1999). Gründe und Vorschläge für eine Allgemeine Psychotherapie [Reasons and proposals for a general psychotherapy]. *Psychotherapeut, 44*(6), 350-359. doi:10.1007/s002780050190

Grawe, K., Caspar, F., & Ambühl, H. (1990). Die Berner Therapievergleichsstudie: Wirkungsvergleich und differentielle Indikation [The Bern therapy comparison study: Comparison of effectiveness and differential indication]. *Zeitschrift für klinische Psychologie, 19*(4), 338-361.

Greenberg, L. S. (2002). Integrating an emotion-focused approach to treatment into psychotherapy integration. *Journal of Psychotherapy Integration, 12*(2), 154-189. doi:10.1037/1053-0479.12.2.154

Greenberg, L. S., & Pinsof, W. M. (1986). *The psychotherapeutic process: A research handbook*. New York: Guilford Press.

Grosse Holtforth, M., & Grawe, K. (2000). Fragebogen zur Analyse motivationaler Schemata (FAMOS) [Questionnaire for the Analysis of Motivational Schemas]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 29*(3), 170-179. doi:10.1026//0084-5345.29.3.170

Grosse Holtforth, M., & Grawe, K. (2003). Der Inkongruenzfragebogen (INK) [The incongruence questionnaire (INK). An instrument for the analysis of motivational incongruence]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 32*(4), 315-323. doi:10.1026/0084-5345.32.4.315

Grosse Holtforth, M., Krieger, T., Zimmermann, J., Altenstein-Yamanaka, D., Dörig, N., Meisch, L., & Hayes, A. M. (2019). A randomized-controlled trial of cognitive–behavioral therapy for depression with integrated techniques from emotion-focused

and exposure therapies. *Psychotherapy Research, 29*(1), 30-44. doi:10.1080/10503307.2017.1397796

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine, 363*(4), 301-304. doi:10.1056/NEJMp1006304

Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy of depression: pretreatment patient predictors of outcome. *Clinical Psychology Review, 22*(6), 875-893. doi:10.1016/s0272-7358(02)00106-x

Hautzinger, M., Joormann, J., & Keller, F. (2005). *DAS: Skala dysfunktionaler Einstellungen; Manual [Scale of dysfunctional attitudes]*. Göttingen: Hogrefe.

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Das Beck Depressionsinventar II. Deutsche Bearbeitung und Handbuch zum BDI-II [Beck Depressive Inventory II. German Revision and Manual for the BDI-II]*. Frankfurt a.M.: Harcourt Test Services.

Hill, C. E., & Corbett, M. M. (1993). A perspective on the history of process and outcome research in counseling psychology. *Journal of Counseling Psychology, 40*(1), 3-24. doi:10.1037/0022-0167.40.1.3

Hiroe, T., Kojima, M., Yamamoto, I., Nojima, S., Kinoshita, Y., Hashimoto, N., . . . Furukawa, T. A. (2005). Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. *Psychiatry Research, 135*(3), 229-235. doi:10.1016/j.psychres.2004.03.014

Hollon, S. D., Cohen, Z. D., Singla, D. R., & Andrews, P. (2019). Recent Developments in the Treatment of Depression. *Behavior Therapy, 50*(2), 257-269. doi:10.1016/j.beth.2019.01.002

Holsboer, F. (2008). How can we realize the promise of personalized antidepressant medicines? *Nature Reviews Neuroscience, 9*(8), 638-646. doi:10.1038/nrn2453

Horvath, A. O., & Greenberg, L. S. (1994). *The working alliance: Theory, research, and practice*. New York: John Wiley & Sons.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*(10), 1059-1064. doi:10.1037//0003-066x.51.10.1059

Huibers, M. J., Cohen, Z. D., Lemmens, L. H., Arntz, A., Peeters, F. P., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PloS One, 10*(11), e0140771. doi:10.1371/journal.pone.0140771

Hupfeld, J., & Ruffieux, N. (2011). Validierung einer deutschen version der Self-Compassion Scale (SCS-D) [Validation of a German version of the Self-Compassion Scale (SCS-D)]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 40*(2), 115-123. doi:10.1026/1616-3443/a000088

Jerusalem, M., & Schwarzer, R. (1999). *Skala zur allgemeinen Selbstwirksamkeitserwartung*. Berlin: Freie Universität Berlin.

Katsnelson, A. (2013). Momentum grows to make 'personalized' medicine more 'precise'. *Nature Medicine, 19*(3), 249. doi:10.1038/nm0313-249

Keefe, J. R., Wiltsey Stirman, S., Cohen, Z. D., DeRubeis, R. J., Smith, B. N., & Resick, P. A. (2018). In rape trauma PTSD, patient characteristics indicate which trauma-focused treatment they are most likely to complete. *Depression and Anxiety, 35*(4), 330-338. doi:10.1002/da.22731

Kenter, R. M., van de Ven, P. M., Cuijpers, P., Koole, G., Niamat, S., Gerrits, R. S., . . . van Straten, A. (2015). Costs and effects of Internet cognitive behavioral treatment

blended with face-to-face treatment: results from a naturalistic study. *Internet Interventions, 2*(1), 77-83. doi:10.1016/j.invent.2015.01.001

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593-602. doi:10.1001/archpsyc.62.6.593

Klaghofer, R., & Brähler, E. (2001). Konstruktion und Teststatistische Prüfung einer Kurzform der SCL-90–R [Construction and test statistical evaluation of a short version of the SCL-90–R]. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie, 49*(2), 115-124.

Kleiboer, A., Smit, J., Bosmans, J., Ruwaard, J., Andersson, G., Topooco, N., . . . Baños, R. (2016). European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. *Trials, 17*(1), 387. doi:10.1186/s13063-016-1511-1

Kolden, G. G. (1991). The generic model of psychotherapy: An empirical investigation of patterns of process and outcome relationships. *Psychotherapy Research, 1*(1), 62-73. doi:10.1080/10503309112331334071

Kolden, G. G., & Howard, K. I. (1992). An empirical test of the generic model of psychotherapy. *The Journal of Psychotherapy Practice and Research, 1*(3), 225-236.

Kooistra, L. C., Ruwaard, J., Wiersma, J. E., van Oppen, P., van der Vaart, R., van Gemert-Pijnen, J. E., & Riper, H. (2016). Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. *Internet Interventions, 4*, 61-71. doi:10.1016/j.invent.2016.01.003

Kooistra, L. C., Wiersma, J. E., Ruwaard, J., van Oppen, P., Smit, F., Lokkerbol, J., . . . Riper, H. (2014). Blended vs. face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost-effectiveness trial. *BMC Psychiatry, 14*(1), 290. doi:10.1186/s12888-014-0290-z

Kraemer, H. C. (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Statistics in Medicine, 32*(11), 1964-1973. doi:10.1002/sim.5734

Kraemer, H. C., & Thiemann, S. (1989). A strategy to use soft data effectively in randomized controlled clinical trials. *Journal of Consulting and Clinical Psychology, 57*(1), 148-154. doi:10.1037//0022-006x.57.1.148

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*(10), 877-883. doi:10.1001/archpsyc.59.10.877

Kramer, U., & Stiles, W. B. (2015). The responsiveness problem in psychotherapy: A review of proposed solutions. *Clinical Psychology: Science and Practice, 22*(3), 277-295. doi:10.1111/cpsp.12107

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. doi:10.1046/j.1525-1497.2001.016009606.x

Kühner, C., Huffziger, S., & Nolen-Hoeksema, S. (2007). *Response Styles Questionnaire-Deutsche Version: RSQ-D; Manual*. Göttingen: Hogrefe.

Lambert, M. J. (2013). *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (M. J. Lambert Ed. 6th ed.). Hoboken, New Jersey: John Wiley & Sons.

Lampropoulos, G. K. (2001). Bridging technical eclecticism and theoretical integration: Assimilative integration. *Journal of Psychotherapy Integration, 11*(1), 5-19. doi:10.1023/A:102667280

Laynard, R., Clark, D., Knapp, M., & Mayraz, G. (2007). Cost-benefit analysis of psychological therapy. *National Institute Economic Review, 202*(1), 90-98. doi:10.1177/0027950107086171

Lemmens, L. H., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huibers, M. J. (2019). Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: follow-up of a randomized controlled trial. *Psychological Medicine, 49*(3), 465-473. doi:10.1017/S0033291718001083

Llewelyn, S., Macdonald, J., & Doorn, K. (2016). Process–outcome studies. In J. C. Norcross, G. R. VandenBos, D. K. Freedheim, & B. O. Olatunji (Eds.), *APA handbooks in psychology®. APA handbook of Clinical Psychology: Theory and research* (pp. 451-463). Washington, DC: American Psychological Association.

Lopez-Gomez, I., Lorenzo-Luaces, L., Chaves, C., Hervas, G., DeRubeis, R. J., & Vázquez, C. (2019). Predicting optimal interventions for clinical depression: Moderators of outcomes in a positive psychological intervention vs. cognitive-behavioral therapy. *General Hospital Psychiatry, 61,* 104-110. doi:10.1016/j.genhosppsych.2019.07.004

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample Size Requirements for Multivariate Models to Predict Between-Patient Differences in Best Treatments of Major Depressive Disorder. *Clinical Psychological Science, 7*(3), 445-461. doi:10.1177/2167702618815466

Lutz, W., Clausen, S. A., & Deisenhofer, A.-K. (2019). Perspektiven einer evidenzbasierten und personalisierten Psychotherapie: Der Trier Therapie Navigator (TTN) [Perspectives of an Evidence-Based and Personalized Psychotherapy: The Trier

Treatment Navigator (TTN)]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 0*(0), 1-11. doi:10.1026/1616-3443/a000518

Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D., . . . Tholen, S. (2006). Empirically and clinically useful decision making in psychotherapy: Differential predictions with treatment response models. *Psychological Assessment, 18*(2), 133-141. doi:10.1037/1040-3590.18.2.133

Lutz, W., Zimmermann, D., Muller, V., Deisenhofer, A. K., & Rubel, J. A. (2017). Randomized controlled trial to evaluate the effects of personalized prediction and adaptation tools on treatment outcome in outpatient psychotherapy: study protocol. *BMC Psychiatry, 17*(1), 306. doi:10.1186/s12888-017-1464-2

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine, 3*(11), e442. doi:10.1371/journal.pmed.0030442

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*: University of Minnesota Press.

Orlinsky, D. E. (2009). The “Generic Model of Psychotherapy” after 25 years: Evolution of a research-based metatheory. *Journal of Psychotherapy Integration, 19*(4), 319-339. doi:10.1037/a0017973

Orlinsky, D. E., & Howard, K. (1986). Process and outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Bergin & Garfield's Handbook of Psychotherapy and Behavior Change* (3rd ed., pp. 311-384). New York: John Wiley & Sons.

Orlinsky, D. E., & Howard, K. I. (1987). A generic model of psychotherapy. *Journal of Integrative & Eclectic Psychotherapy, 6*(1), 6-27.

Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. *Bergin and Garfield’s handbook of psychotherapy and behavior change, 5*, 307-389.

Orlinsky, D. E., & Russell, R. L. (1994). Tradition and change in psychotherapy research. In R. L. Russell (Ed.), *Reassessing psychotherapy research*. New York: Guilford Press.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*(2), 109-118. doi:10.1037/h0024436

Penninx, B. W., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P., . . . van der Meer, K. (2011). Two-year course of depressive and anxiety disorders: results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders, 133*(1-2), 76-85. doi:10.1016/j.jad.2011.03.027

Perlis, R., Iosifescu, D., Castro, V., Murphy, S., Gainer, V., Minnier, J., . . . Gallagher, P. (2012). Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychological Medicine, 42*(1), 41-50. doi:10.1017/S0033291711000997

Petkova, E., Ogden, R. T., Tarpey, T., Ciarleglio, A., Jiang, B., Su, Z., . . . Trivedi, M. H. (2017). Statistical analysis plan for stage 1 EMBARC (Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care) study. *Contemporary Clinical Trials Communications, 6*, 22-30. doi:10.1016/j.conctc.2017.02.007

Rice, L. N., & Greenberg, L. S. (1984). *Patterns of change: Intensive analysis of psychotherapy process*. New York: Guilford Press.

Roth, M., Decker, O., Herzberg, P. Y., & Brähler, E. (2008). Dimensionality and norms of the Rosenberg Self-Esteem Scale in a German general population sample. *European Journal of Psychological Assessment, 24*(3), 190-197. doi:10.1027/1015-5759.24.3.190

Röthlin, P., Holtforth, M. G., Bergomi, C., Berking, M., Ottenbreit, N. D., & Caspar, F. (2010). Vermeidung und Depression [Avoidance and depression]. *Diagnostica, 56*(1), 46-55. doi:10.1026/0012-1924/a000008

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., . . . Lebowitz, B. D. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. *American Journal of Psychiatry, 163*(11), 1905-1917. doi:10.1176/ajp.2006.163.11.1905

Schulte, D. (2005). Messung der Therapieerwartung und Therapieevaluation von Patienten (PATHEV) [Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV)]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 34*(3), 176-187. doi:10.1026/1616-3443.34.3.176

Schwaederle, M., Zhao, M., Lee, J. J., Eggermont, A. M., Schilsky, R. L., Mendelsohn, J., . . . Kurzrock, R. (2015). Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. *Journal of Clinical Oncology, 33*(32), 3817-3825. doi:10.1200/JCO.2015.61.599

Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: can we match patients with treatments? *American Journal of Psychiatry, 167*(12), 1445-1455. doi:10.1176/appi.ajp.2010.09111680

Smith, M., Glass, G., & Miller, T. (1980). Meta-analysis of psychotherapy. *American Psychologist, 41,* 165-180.

Spek, V., Nyklíček, I., Cuijpers, P., & Pop, V. (2008). Predictors of outcome of group and internet-based cognitive behavior therapy. *Journal of Affective Disorders, 105*(1), 137-145. doi:10.1016/j.jad.2007.05.001

Stekhoven, D. J. (2013). missForest: Nonparametric Missing Value Imputation using Random Forest. R package version, 1, 4.

Stiles, W. B. (2009). Responsiveness as an Obstacle for Psychotherapy Outcome Research: It's Worse Than You Think. *Clinical Psychology: Science and Practice, 16*(1), 86-91. doi:10.1111/j.1468-2850.2009.01148.x

Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice, 5*(4), 439-458. doi:10.1111/j.1468-2850.1998.tb00166.x

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(4), 547-559. doi:10.1037/a0028226

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894-904. doi:10.1037/0022-006X.67.6.894

Thase, M. E., Reynolds, C. F., Frank, E., & Simons, A. D. (1994). Response to cognitive-behavioral therapy in chronic depression. *Journal of Psychotherapy Practice & Research, 3*(3), 304-314.

Thomas, A., Brähler, E., & Strauß, B. (2011). IIP-32: Entwicklung, Validierung und Normierung einer Kurzform des Inventars zur Erfassung interpersonaler Probleme [IIP-32: Development, validation, and standardization of a short form of the Inventory of Interpersonal Problems]. *Diagnostica, 57*(2), 68-83. doi:10.1026/0012-1924/a000034

Timulak, L. (2008). Process-outcome research. In *Research in Psychotherapy and Counselling* (pp. pages 127-133). London: SAGE Publications Ltd.

Trösken, A., & Grawe, K. (2002). *Das Berner Ressourceninventar: Konstruktion und Validierung zweier Instrumente zur Erfassung von Patientenressourcen für die klinische Forschung und Praxis. [The Bernese Resource Inventory: construction and validation of two instruments for measuring patient resources for clinical research and practice]*. Unpublished doctoral thesis. Department of Psychology. University of Bern. Bern, Switzerland.

van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L. H., Peeters, F. P., Keefe, J. R., Cohen, Z. D., & Huibers, M. J. (2019). Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: cognitive therapy or interpersonal psychotherapy? *Psychological Medicine*, 1-11. doi:10.1017/S0033291719003192

van Bronswijk, S. C., Lemmens, L. H., Keefe, J. R., Huibers, M. J., DeRubeis, R. J., & Peeters, F. P. (2018). A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder. *Depression and Anxiety, 36*(3), 252-261. doi:10.1002/da.22868

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., . . . Parsey, R. (2019). Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychological Medicine, 49*(7), 1118-1127. doi:10.1017/S0033291718001708

Westermann, S., grosse Holtforth, M., & Michalak, J. (2019, in press). Motivation in Psychotherapy. In R. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (2nd ed., pp. 417-441). New York: Oxford University Press.

World Health Organization. (1995). The World Health Organization Quality of Life Assessment (WHO-QOL): Position paper from the World Health Organization. *Social Science & Medicine, 41*(10), 1403-1409. doi:10.1016/0277-9536(95)00112-K

World Health Organization. (2008). The global burden of disease: 2004 update. Retrieved from https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/

World Health Organization. (2018). Depression. Retrieved from https://www.who.int/news-room/fact-sheets/detail/depression

Zilcha-Mano, S. (2018). Major developments in methods addressing for whom psychotherapy may work and why. *Psychotherapy Research, 29*(6), 693-708. doi:10.1080/10503307.2018.1429691

Zilcha-Mano, S., Keefe, J. R., Chui, H., Rubin, A., Barrett, M. S., & Barber, J. P. (2016). Reducing Dropout in Treatment for Depression: Translating Dropout Predictors Into Individualized Treatment Recommendations. *Journal of Clinical Psychiatry, 77*(12), e1584-e1590. doi:10.4088/JCP.15m10081

5 Declaration of Originality

Philosophisch-humanwissenschaftliche Fakultät
Dekanat
Fabrikstrasse 8, CH-3012 Bern

UNIVERSITÄT BERN

Erklärung zur Dissertation

Hiermit bestätige ich, dass ich die Dissertation (Titel):

im Fach

unter der Leitung von Prof. Dr.

ohne unerlaubte Hilfe ausgeführt und an keiner anderen Universität zur Erlangung eines akademischen Grades eingereicht habe.

Datum Unterschrift